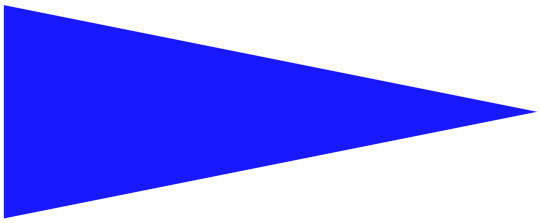
About several Reference Processings in Multimodal Human-Computer Communication

Ali Choumane, Jacques Siroux

I R I S A
INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTÈMES ALÉATOIRES

P U B L I C A T I O N
I N T E R N E
N° 1845



ABOUT SEVERAL REFERENCE PROCESSINGS IN MULTIMODAL HUMAN-COMPUTER COMMUNICATION

ALI CHOUMANE AND JACQUES SIROUX

ISSN 1166-8687

IRISA
CAMPUS UNIVERSITAIRE DE BEAULIEU - 35042 RENNES CEDEX - FRANCE

IRISA

INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTÈMES ALÉATOIRES
Campus de Beaulieu – 35042 Rennes Cedex – France
Tél. : (33) 02 99 84 71 00 – Fax : (33) 02 99 84 71 71
http://www.irisa.fr

About several Reference Processings in Multimodal Human-Computer Communication

Ali Choumane * and Jacques Siroux **

Systèmes cognitifs
Projets Cordial

Publication interne n1845 — Avril 2007 — 67 pages

Abstract: We are interested in multimodal systems that use the following modes and modalities : speech (and natural language) as input as well as output, gesture as input and visual as output using screen displays. The user exchanges with the system by gesture and/or oral statements in natural language. This exchange, encoded in the different modalities, carries the goal of the user and also the designation of objects (referents) needed to achieve this goal. The system must identify in a precise and non-ambiguous way the objects designated by the user. In this document, we expose a state of the art about the processing of referential activities in this particular context. We address methods and algorithms of processing of linguistic and multimodal referential activities. We also address problems relied to some designations by gesture. Our point of view is to situate studies and results within the framework aiming to propose generic solutions and to develop the GEORAL multimodal system.

Key-words: Multimodal human-computer communication, reference, anaphora, definite description, designation by gesture

(Résumé : tsvp)

Ces travaux sont partiellement financés par le contrat 211-B2-9/ARED 1800 du conseil régional de Bretagne.

* ali.choumane@irisa.fr
** jacques.siroux@univ-rennes1.fr

Centre National de la Recherche Scientifique
(UMR 6074) Université de Rennes 1 – Insa de Rennes

Institut National de Recherche en Informatique
et en Automatique – unité de recherche de Rennes

Sur quelques traitements de la référence dans le cadre de communication homme-machine multimodale

Résumé : Nous nous intéressons aux systèmes multimodaux qui utilisent les modes et modalités suivantes : l'oral (et le langage naturel) en entrée et en sortie, le geste en entrée et le visuel en sortie par affichage sur écran. L'usager échange avec le système par un geste et/ou un énoncé oral en langue naturelle. Dans cet échange, encodé sur les différentes modalités, se trouvent l'expression du but de l'usager et la désignation des objets (référents) nécessaires à la réalisation de ce but. Le système doit identifier de manière précise et non ambiguë les objets désignés par l'usager. Nous exposons dans ce document un état de l'art sur le traitement des activités référentielles dans ce contexte particulier. Nous abordons des méthodes et algorithmes de traitement des activités référentielles linguistiques et multimodales ainsi que les problèmes posés par certains gestes désignation. Notre point de vue est de positionner les études et résultats dans le cadre de travaux visant à proposer des solutions génériques et à développer le système multimodal GEORAL.

Mots clés : Communication homme-machine multimodale, référence, anaphore, description définie, désignation par geste

1 Cadre de l'étude

1.1 Communication personne-machine multimodale

Cette étude se situe dans le contexte des systèmes de communication personne-machine multimodaux.

Le but de tels systèmes est de permettre aux utilisateurs d'obtenir la réalisation de services. Par exemple actuellement des systèmes multimodaux sont conçus pour fournir des renseignements sur des horaires de vols aériens, pour élaborer des itinéraires ou encore pour aider la réalisation de maquettes ou de plans. Le service est rendu par une partie du système désignée souvent sous le terme d'*application*. L'application, comme le suggère les exemples ci-dessus, peut nécessiter le recueil d'un nombre variable d'informations auprès de l'usager et peut mettre à la disposition de l'usager plusieurs services (les tâches) réalisables en parallèle ou en séquence.

Le terme *multimodal* fait référence au fait que différents modes de communication sont utilisés durant les interactions entre utilisateur et système, et de manière indirecte aux structures des informations rattachées au mode (les modalités de communication).

Les premiers systèmes mis au point étaient pratiquement monomodaux avec d'abord l'usage de l'écrit (clavier) puis rapidement avec l'utilisation du mode oral et d'une modalité langagière très restreinte (mots clés, langages pseudo-naturels). Le mode oral et la modalité langagière se sont en effet imposés comme étant particulièrement intéressants parce que chercheurs et industriels ont estimé (et estiment toujours) que ce type de communication est naturel et que son usage n'est pas soumis à un apprentissage préalable ; on pense ainsi rendre possible à tous types d'utilisateur l'accès aux systèmes. Cependant les difficultés posées par la compréhension automatique de la parole ont amené les chercheurs à envisager l'utilisation d'autres modes et les progrès techniques ainsi que les avancées scientifiques ont permis d'évoluer naturellement des systèmes monomodaux vers des systèmes multimodaux avec toujours l'utilisation du mode oral, mais avec une modalité langagière élaborée, associé à d'autres modes de communication tel que le toucher. Cette évolution n'est bien sûr pas terminée car de nombreux obstacles techniques restent à lever.

En prenant en compte leur objectif de service, de tels systèmes peuvent être considérés comme de véritables intermédiaires entre les utilisateurs humains et l'application [19], intermédiaires dont le rôle est de mener un double dialogue (utilisateur-dialogueur, dialogueur-application). Ce double dialogue est nécessaire pour faire émerger un univers virtuel (ou réel pour certaines applications visuelles) connu et accepté de l'utilisateur et de l'application. Il s'agit d'atteindre un consensus entre buts de l'utilisateur et possibilités de l'application. Cet accord englobe une compréhension mutuelle des intentions qui peuvent apparaître (et qu'il faut satisfaire) durant les différentes phases de l'interaction mais aussi une vue partagée sur toutes les entités (paramètres, objets, ...) manipulées et nécessaires à l'accomplissement de la

tâche. L'utilisateur désigne ces entités en recourant aux modes et modalités à sa disposition : oral, langue, geste... On parle d'activités référentielles. Un rôle important d'un système est de reconnaître et de comprendre ces activités référentielles.

Ces activités de reconnaissance, de compréhension et d'interprétation est compliqué car le système est confronté à de nombreuses difficultés. D'une part, la performance de l'utilisateur dans l'activité de désignation n'est pas sûre, elle peut être entachée d'ambiguïtés, d'erreurs, d'hésitation et conduit à des « bruits » ou des malentendus qui sont susceptibles d'être aggravés par les dispositifs matériels et les programmes du système. D'autre part, les activités de désignation se déroulent durant la constitution de l'espace virtuel commun : les partenaires procèdent souvent par ajustements : les quiproquo et ambiguïtés sont usuels et fréquents. Enfin, bien que la multimodalité soit normalement utilisée pour améliorer la communication et diminuer le nombre d'ambiguïtés, l'usage conjoint de plusieurs modes multiplie les problèmes techniques et peut dégrader les performances des usagers.

Le propos de ce document est d'exposer un état de l'art sur le traitement de ces activités référentielles dans ce contexte particulier. Trois points de vue seront plus particulièrement pris en considération :

- la modalité langagière et la linguistique : des fragments d'énoncés des utilisateurs servent notamment à désigner des entités ou encore à parler d'entités déjà désignées : on parle d'expressions référentielles (ER). De nombreuses études linguistiques ont été réalisées sur le sujet et diverses théories ont été élaborées. Cependant, peu de ces théories ont été implémentées dans des systèmes mais elles demeurent néanmoins fondamentales.
- la modalité langagière et le traitement automatique de la langue naturelle : des algorithmes intéressants fondés pour certains sur des résultats de recherche linguistique ont été proposés dans le cadre de systèmes de dialogue ou dans de le cadre d'application en langue naturelle (fouille de données par exemple)
- la modalité haptique : les problèmes particuliers de l'utilisation du geste de désignation conjointement avec des ER seront abordés.

Nous n'aborderons les travaux qui portent sur l'intentionnalité et les modalités de désignation [10] [9][21]. Bien que leurs résultats soient très intéressants, ces types de travaux nous semblent se situer à un niveau qui ne correspond complétement pas ni au type de système visé ni à la mise en œuvre projetée.

1.2 GEORAL application support

Le projet IRISA/CORDIAL travaille dans le domaine des systèmes multimodaux à forte composante orale. Nous avons mis au point le système GEORAL tactile dont l'évolution suit celle décrite plus haut : nous disposons actuellement d'un système multimodal : oral, visuel, gestuel alors que la première version du système était seulement à composante principale orale. Au fur et à mesure de l'évolution du système et de l'augmentation du nombre de ses fonctionnalités, des problèmes pour traiter les activités référentielles des usagers sont

apparus. Avec l'aide de la région Bretagne, nous avons proposé le projet REPAIMTA[1] dont l'objectif est d'élaborer un modèle qui permet un traitement générique des activités référentielles dans le cadre de systèmes multimodaux. Cette section a pour objet une présentation sommaire du système GEORAL tactile et de son environnement, présentation qui permettra d'illustrer certains aspects de la section précédente et qui servira de support à des exemples dans la suite du document.

1.2.1 Description

Géoral est un système multimodal pour une application de renseignements touristiques sur le Trégor (région de Bretagne). L'usager peut demander des informations et la localisation de sites touristiques comme plage, église, camping en précisant un endroit , une zone (dessinée ou située par rapport à un élément géographique ou cartographique particulier : rivière, route, côte) ; il peut également demander la distance et l'itinéraire entre deux localités.

Les modes et modalités mis à la disposition de l'usager sont les suivants :
- l'oral en entrée et en sortie du système. L'usager peut formuler ses demandes ou ses réponses aux questions du système à la voix et en langage naturel de manière spontanée (pas de consignes particulières d'élocution). Certaines réactions du système sont aussi transmises oralement à l'utilisateur.
- le mode visuel : le système affiche sur un écran une carte de la région ; cette carte contient des informations géographiques et touristiques habituelles : routes, rivières, fleuves. Des effets de zoom, de surlignage, de clignotement permettent au système de focaliser l'attention de l'usager.
- le mode gestuel par l'intermédiaire d'un écran tactile : l'usager peut désigner par différents types de geste des éléments sur la présentation affichée à l'écran.

1.2.2 Traitements et dispositifs matériel et logiciel

Le traitement d'un échange entre l'utilisateur et le système passe par les étapes suivantes :
- La première étape concerne le traitement du signal vocal et la reconnaissance de la parole. Nous utilisons actuellement le moteur de reconnaissance de la parole Philsoft ©V3 développé et distribué par TELISMA dans un contexte de parole continue spontanée (présence de phénomènes velléitaires). Il faut noter que le reconnaisseur peut faire des erreurs de reconnaissance (omission, substitution, insertion de mot) qui sont combinées avec les problèmes de performance de l'usager.
- La deuxième étape est la compréhension de l'énoncé de l'usager : analyse syntaxique et début d'analyse sémantique. L'analyseur syntaxique utilisé est relativement simple (algorithme par « différence de listes » [13]). Parallèlement à la reconnaissance de parole et à l'étape de compréhension, les activités tactiles de l'usager sont traitées : reconnaissance et reconstitution du geste et début d'interprétation pragmatique en

[1] **RÉ**férence **PA**role **IM**age **TA**ctile, contrat 211-B2-9/ARED 1800

fonction de l'affichage présent sur l'écran. Ici aussi des problèmes dûs à la performance de l'usager peuvent survenir : hésitations, corrections, imprécisions, ...
- La troisième étape concerne la fusion des modalités : il est nécessaire de construire un acte de communication complet du point de vue référentiel en prenant en compte les résultats de l'analyse syntactico-sémantique et celui de l'interprétation de l'activité tactile. Le modèle utilisé pour cette fusion est fondé sur une modélisation des actes communicatifs à l'aide d'opérateurs de plan. Durant ce processus, nous prenons en compte certaines erreurs ou incompatibilités entre modes qui peuvent être traitées en fonction de l'historique de dialogue ou en fonction de connaissances que le système peut posséder sur l'utilisateur.
- L'étape suivante concerne l'interprétation dialogique de l'acte communicatif complet ; cette interprétation peut nécessiter un accès à la base de données géographiques. Cette étape met en œuvre un modèle de dialogue de type automate à états fini. Nous utilisons également des algorithmes de relâchement de contraintes en cas de réponse vide pour permettre au système de relancer l'interaction [18] de manière coopérative.
- La dernière étape produit la réponse du système sous forme d'affichage de la carte ou d'une partie de la carte (zoom, clignotement) et sous forme orale (synthèse de la parole). Nous utilisons le moteur de synthèse de la parole fourni par FT R&D.

La mise en œuvre de tous ces processus est effectuée sur la plate forme multiagent DORIS [28]. Son implémentation est fondée sur le logiciel médiateur (middleware) JADE[2] qui est distribué par TILab en Open Source. JADE respecte les standards de la FIPA[3] et permet de simplifier l'implémentation des systèmes multiagents. DORIS bénéficie de la technologie client/serveur avec le navigateur Web standard pour le client et HTTP/applet avec JSP du côté serveur. La puissance de cette architecture, d'un point de vue client, est la large diffusion de machine JAVA avec J2SE pour les systèmes desktops ou J2ME pour les systèmes embarqués comme les téléphones mobiles et PDAs. L'un des avantages de cette implémentation sous JADE est sa flexibilité : il est possible de modifier très facilement les protocoles de coopération entre agents, le remplacement d'un agent par un autre agent (changement d'algorithme par exemple) est relativement aisé grâce au concept d'autonomie. Tout cela permet une rapidité de mise à jour du système en fonction de l'évolution des techniques ou de l'application.

Chaque modalité sous DORIS est liée à un agent-proxy JADE. Cet agent échange des messages avec ses collègues via la plate forme agent. L'accès aux ressources de synthèse et de reconnaissance de la parole s'effectue via le protocole MRCP[4]. Le protocole est chargé de gérer l'acheminement des message entre le client et l'agent-proxy dédié à la parole.

[2] Java Agent DEvelopment Framework.
[3] Foundation for Intelligent Physical Agents, une association de normalisation dont le but est de fournir des standards pour faciliter l'interopérabilité entre les systèmes à base d'agents.
[4] Media Ressource Control Protocol.

Les informations interrogées (base de données) sont actuellement dans une base de faits PROLOG. Pour l'instant le lien entre les faits et l'affichage est réalisé de manière ad hoc. Nous cherchons à l'améliorer en concevant une base de données véritablement géographique avec affichage vectoriel en SVG[5]. D'une part cela permettra de réduire le nombre de requêtes vers l'agent Base de données, et d'autre part des fonctionnalités correspondent aux nécessités du projet (zoom, déplacement, détection instantanée des zones et des polylignes ...) pourront être mises en œuvre plus facilement. En effet, avec SVG, les coordonnées, dimensions et structures des objets vectoriels (ou géographiques) sont indiquées sous forme numérique dans un document XML. Le format permet l'intégration d'animations ou la manipulation d'objets graphiques par programmation, notamment grâce à des scripts qui peuvent être intégrés dans le SVG. Un des intérêts majeurs de SVG est qu'il peut être inclus dans d'autres documents XML, comme par exemple des documents XHTML ou des documents XML devant être traités par XSL-FO. Respectant le norme XML, une image SVG peut également être manipulée par l'intermédiaire du modèle DOM (Document Object Model). SVG peut être visualisé nativement avec certains navigateurs Web, ce qui convient à la portabilité de l'interface d'accès à Géoral et surtout la perspective d'accès de cette dernière à partir d'un PDA.

1.3 Exemples commentés de dialogue

Nous présentons quelques extraits de dialogue avec le système actuel ainsi que des dialogues que nous voudrions voir se réaliser. (U : usager, S : système, en italique : observations ou commentaires et activités autres que langagières). Certains termes de vocabulaire sont explicités plus loin.

Exemple 1 *L'utilisateur formule sa requête initiale uniquement par oral.*
U1 : je voudrais les plages à Perros-Guirec.
Le système va effectuer une demande de confirmation implicite en répétant le nom de la localité.
S1 : Je recherche les plages à Perros-Guirec, veuillez patienter + *clignotement sur le nom de la ville*
S2 : *effet de zoom, affichage d'une icône et par la synthèse de la parole production de l'énoncé* : plage de Trestignel, Trestignel (*énonciation du nom de site et de sa localisation)*
U2 : silence (grounding)
Le silence est pris comme approbation, le système continue à énumérer des éléments de sa réponse. S3 : *affichage d'une icône, avec la synthèse de la parole* : plage de Trestraou, Trestraou
L'utilisateur initialise un sous dialogue
U3 :Pouvez-vous l'écrire ?
S4 : *affichage d'une icône sur l'endroit, avec la synthèse de la parole* : plage de Trestraou,

[5]Scalable Vector Graphics.

Trestraou, *incrustation à l'écran de la chaîne de caractères*
U4 : silence
S5 : voulez-vous un autre renseignement ?

Exemple 2 *L'utilisateur formule la requête initiale uniquement par oral, des problèmes de reconnaissance apparaissent, il n'y pas de réponse dans la base de données ; on note la présence d'une anaphore dans l'énoncé U5*
U1 : je voudrais les campings à Trégastel.
S1 : *le nom Trégastel n'est pas reconnu.* je recherche les campings ; dans quelle localité voulez-vous la recherche ?
U2 : à Trégastel
S2 : Trégastel, veuillez patienter.
S3 : *il n'y a pas de réponse dans la base, relâchement de contraintes et production d'une réponse suggestive* il n'y a pas de campings à Trégastel, mais il y en a dans les localités suivantes : Perros-Guirec, Lannion , *clignotement sur les localités*, dans laquelle voulez-vous effectuer la recherche ?
U3 : à Lannion
S4 : camping de Beg Leguer, Beg Leguer, *zoom et affichage d'une icône sur l'endroit*
U4 : silence
S5 : camping des deux rives, Lannion, *zoom et affichage d'une icône sur l'endroit*
U5 : vous pouvez écrire le nom du premier

Exemple 3 L*'utilisateur formule la requête initiale uniquement par oral, la requête concerne une demande de trajet.*
U1 : je voudrais la distance entre Lannion et Tréguier.

Exemple 4 *La requête initiale est multimodale, les modalités orale et tactile sont utilisées de manière complétive.*
U1 : Je voudrais les campings *ici + pointé sur l'écran montrant l'endroit visé.* S1 : je recherche les campings à Lannion....

Exemple 5 *La requête initiale est multimodale, les modalités orale et tactile sont utilisées de manière complétive.*
U1 : Je voudrai les campings *ici + dessin d'une zone sur l'écran montrant l'endroit visé.*
S1 : Je recherche les campings autour de Trégastel.

Exemple 6 *La requête initiale est multimodale, les modalités orale et tactile sont utilisées, la modalité tactile est utilisée de manière supplétive.*
U1 : Je voudrais les campings *+ dessin d'une zone sur l'écran montrant l'endroit visé.*

Exemple 7 *Requête multimodale : les modalités orale et tactile sont utilisées de manière confirmative.*
U1 : Je voudrais les campings à Lannion *+ pointage sur l'écran montrant la ville de Lannion.*

Exemple 8 *Une requête multimodale future* :
U1 : je voudrais les campings le long des berges de cette rivière + geste de tracé d'une ligne le long d'un élément longiligne sur la carte.
Le trait peut ne pas désigner une rivière, le tracé peut aussi porter à confusion s'il est près des représentations d'une route et d'une rivière voisines. La reconnaissance de la parole peut aussi entraîner des problèmes : par exemple reconnaissance de lisière au lieu de rivière. Notons aussi que les berges ne sont pas explicitement représentées sur la carte affichée.

Exemple 9 L*a requête initiale est multimodale, l'élément nécessaire à la recherche n'est pas présent sur la carte ni dans la base de données. L'énoncé Un montre une référence anaphorique graphique.*
U1 : je voudrais les campings dans ce triangle + 3 pointés
...
Un : et sur le sommet

2 Désignation et activités référentielles

Le but du dialogue dans les systèmes qui nous concernent a pour objet final l'exécution d'une tâche. Avant son exécution une tâche nécessite que système, application et utilisateur soient arrivés à un accord sur les entités concernées. Par exemple dans Géoral, il est nécessaire que le type de la requête et ses paramètres aient été prononcés (ou montrés pour les paramètres) par l'utilisateur, compris par le système et acceptés par la base de données (exemple 1).

Selon le système et la tâche, la nature de ces entités peut être très diverse : informationnelle concrète ou abstraite, physique, ... Une bonne partie des échanges du dialogue va être consacrée à obtenir cet accord sur les entités : l'utilisateur par l'intermédiaire d'actes que nous appellerons référentiels (ou de désignation), qui peuvent être spécifiques ou combinés à d'autres types d'actions va mentionner (désigner) les entités qui l'intéressent ; le système va de manière réciproque tenter d'identifier les entités désignées, directement ou via des échanges ; il va aussi éventuellement présenter à l'utilisateur de manière explicite les entités qui lui semblent intéressantes pour la tâche à réaliser (exemple 2, S3).

D'autre part, le déroulement d'une tâche pouvant s'étendre dans le temps, il peut être nécessaire pour l'utilisateur ou le système soit de faire ressortir les entités particulièrement importantes ou pertinentes à un moment donné (focus ou foyer d'attention), soit de rappeler l'existence d'une entité déjà évoquée ou d'y faire référence.

Nous nous intéresserons ici plus particulièrement aux actes référentiels qui sont supportés par le mode oral et le mode haptique (par l'intermédiaire d'un écran tactile). La modalité associée à l'oral sera le langage naturel ; pour le mode haptique, des gestes simples seront pris en compte (toucher, dessin de zone, ...). Cependant dans le contexte multimodal de

communication homme-machine, les usagers lors de la performance de leurs actes référentiels utilisent de manière consciente ou inconsciente d'autres modes (en particulier la vision) ainsi que des connaissances linguistiques, encyclopédiques, sur le contexte sémantique et sur l'état courant de l'interaction. Cet usage conjoint de modes, modalités et connaissances engendre la complexité du traitement automatique de reconnaissance et d'interprétation des actes et explique aussi la grande diversité des travaux sur le traitement des actes référentiels. Il justifie la présentation de certaines de ces approches dans le document.

Nous commençons par poser définitions et vocabulaire. Ces éléments sont issus d'horizons très divers : linguistique « traditionnelle » (syntaxe, sémantique et pragmatique), TALN et informatique. Certains des matériaux objets des études (textes écrits par exemple) ou buts poursuivis (caractérisation des usages, résumé de texte par exemple) peuvent paraître éloignés de nos préoccupations (langue parlée en contexte de dialogue, interprétation des activités des usagers) mais il n'en reste pas moins vrai que grand nombre d'observations et résultats constituent un apport précieux pour notre propos. Nous exposons ensuite quelques méthodes et algorithmes pour la résolution des expressions référentielles.

2.1 Définitions

Référence : la référence est la fonction par laquelle une expression de la langue (appelée en général expression référentielle) en emploi dans un énoncé renvoie à une entité du monde extra-linguistique réel ou mental. Cette entité peut être de tout type.

Référent : on appelle référent ce à quoi un signe renvoie dans la réalité extra-linguistique.

Déixis : tout énoncé se réalise dans une situation que définissent des coordonnées spatio-temporelles. Les références à cette situation ou à des entités de cette situation forment la déixis et les éléments linguistiques qui consistent à situer l'énoncé (à l'*embrayer*) sur la situation sont des **déictiques**.

Contextes : L'étude de la référence est une des préoccupations de la pragmatique que certains définissent comme la prise en compte du cadre dans lequel un énoncé est émis. Dans les définitions précédentes, il est ainsi fait allusion aux notions de monde et de situation. Cette notion d'*extériorité* (par rapport à un énoncé proprement dit) a été affinée sous la dénomination de *contexte* (que l'on retrouve en sociologie, informatique, etc.) et a donné lieu à plusieurs propositions de définitions. Ainsi, Bunt [6] classe les éléments pertinents qui interviennent dans une interaction ou dans le traitement d'un énoncé d'une interaction dans cinq catégories de contexte : linguistique, sémantique, physique, social et cognitif. De plus Bunt introduit pour chaque contexte un aspect diachronique (dynamique) en distinguant les éléments présents avant l'interaction proprement dite (appelés aspects globaux) avec une valeur donnée et les éléments nouveaux ou modifiés durant l'interaction (appelés aspects locaux ou dynamiques).

- contexte linguistique : il comprend toutes les propriétés du matériau linguistique pris dans une acception très large. Par exemple, on y trouvera pour l'oral la prosodie, pour l'écrit la ponctuation, la casse, les fontes. Ce type de contexte est aussi appelé (D. Estival [12]) le contexte de l'énonciation. Les aspects globaux incluent, outre les connaissances linguistiques communes, les conventions de l'écrit ou de l'oral, les éventuelles interactions précédentes (constitution d'un langage commun). Les aspects locaux comprennent le matériau linguistique produit durant l'interaction ; c'est ce que l'on appelle aussi le **co-texte**.
- contexte sémantique : il est formé par les éléments relatifs à l'application et aux tâches. L'aspect global couvre l'application en général : quels sont les objets manipulés, leurs caractéristiques (de là vient le terme sémantique) et les liens qu'ils entretiennent. Les aspects locaux regroupent des éléments spécifiques comme l'état de la tâche à un instant donné (par exemple quels sont les objets particuliers qui ont déjà été cités, quels sont leurs statuts : en discussion, confirmé...).
- contexte physique : il est caractérisé par l'environnement, le moment de la communication et les modes de communication utilisables. Par exemple : les interactants se voient-ils ? Les aspects dynamiques concernent l'état de l'environnement et celui des modes de communication à chaque instant.
- contexte social : il couvre le type de situation interactive (prise de rendez-vous dans un cabinet médical, consultation, guichet de renseignements, cours d'enseignement) et les rôles des participants dans cette situation (renseigneur-renseigné, docteur-consultant, enseignant-enseigné). Le contexte institutionnel est un aspect global de ce contexte ainsi que le statut social des intervenants. Les aspects locaux ont trait aux obligations et aux droits des interactants : obligation de répondre, possibilité de prendre l'initiative en fonction du co-texte et du contexte sémantique local.
- contexte cognitif : il comprend les attitudes, les intentions, les connaissances et les croyances des intervenants. En début d'interaction, les intervenants ont des objectifs à réaliser, possèdent des connaissances et des croyances sur la tâche à réaliser (mode de fonctionnement, paramètres nécessaires) et sur l'autre intervenant (son rôle, ses connaissances et capacités). Au fur et à mesure du déroulement de l'interaction, l'ensemble de ces éléments va évoluer : acquisition de connaissances, croyances et savoir faire de chaque intervenant. Un système intelligent se doit d'encoder un minimum de ces éléments.

Les intérêts de cette classification sont d'une part d'énumérer tous les constituants intéressants à prendre en compte dans l'élaboration d'un système et d'autre part de proposer un début de solution pour la réalisation de système. Cependant, comme noté dans [37], la répartition des éléments pertinents entre les différents types de contexte n'est pas si évidente.

Première description des expressions référentielles

M. Guyomard [17] considère les expressions référentielles linguistiques sous deux aspects : la forme et la fonction.

Sur l'axe de la forme, on distingue :

- le nombre : singulier/pluriel,
- le genre : masculin/féminin,
- la structure : réfléchie/non réfléchie (lui, lui-même),
- le type syntaxico-lexical qui comprend les catégories suivantes :
  - les mots dits « indexicaux » : les pronoms (personnel, possessif, démonstratif, relatif), les adverbes de lieu et de temps, certains aspects flexionnels des verbes (passé, présent, futur),
  - les noms propres,
  - les groupes (ou syntagmes) nominaux construits autour d'un nom.

Sur l'axe fonctionnel, on distingue :

- le statut : une expression peut être indéfinie ou définie :
  - indéfinie : on fait référence à une classe d'entité plutôt qu'à une entité particulière (je veux *un* camping),
  - définie (on parle de description définie) : on fait référence à une entité précise et unique (à quelle heure part *le* vol AF312 du 23 mars 2006 ?),
- l'usage qui peut être :
  - introductif : l'ER permet d'introduire un nouvel élément dans le co-texte et dans les contextes sémantique et cognitif commun,
  - anaphorique : l'ER permet de désigner une entité déjà citée ou connue.
    L'**anaphore** se définit comme étant la reprise d'une entité déjà évoquée précédemment dans le co-texte ou connue dans les contextes linguistique, sémantique ou cognitif globaux,
  - déictique : l'ER permet de désigner directement des entités du contexte physique.

La détermination de l'usage (anaphorique, introductif, ...) d'une ER n'est pas simple et fait l'objet d'études et de débats. Elle ne dépend pas en effet automatiquement de la forme et les termes de « *évoquée précédemment* » sont assez vagues. Ainsi, le pronom *il* n'est pas toujours anaphorique, la présence d'un déterminant défini dans un premier énoncé d'un texte ou d'un dialogue peut traduire quelque fois un usage introductif et non anaphorique, la présence d'un démonstratif n'implique pas toujours un usage anaphorique ni déictique. Nous présentons plus loin quelques précisions supplémentaires sur ces notions de forme et d'usage.

Antécédent d'une expression référentielle : c'est une expression linguistique déjà présente dans le co-texte qui a le même référent qu'une expression référentielle donnée et qui peut avoir une forme différente. Toute expression référentielle a un référent mais n'a pas forcément un antécédent (par exemple un pronom personnel utilisé de manière déictique).

Co-référence : deux expressions référentielles sont dites co-référentes si elles ont même référent. Il est possible de voir s'établir dans un texte ou durant un dialogue des *chaînes* de co-référence.

Résolution d'une expression référentielle : c'est le processus qui permet de déterminer le référent de l'expression. Pour un certain nombre d'expressions référentielles et de

méthodes de résolution et par abus de langage, c'est aussi souvent le processus qui permet de trouver l'antécédent (et donc par « transitivité » le référent).

Anaphore graphique : dans un environnement multimodal, il peut exister le cas d'une anaphore qui réfère à un objet non évoqué linguistiquement dans le discours mais qui est apparu dans une des activités physiques lors des échanges précédents. L'antécédent dans ce cas se trouve dans l'historique de ces activités physiques et non dans le seul co-texte. Dans notre application, les activités physiques concernent les affichages de carte à l'écran ou bien les activités de désignation tactile de l'usager, activités dont la rétroaction du système laisse des traces sur l'écran. Pour concrétiser ce phénomène, prenons le cas de dialogue suivant :

Exemple 10
U1 : Je voudrais les campings à Trébeurden.
S1 : *Affichage des campings ...*
U2 : Je voudrais les églises.
S2 : *Affichage des églises ...*
U3 : Je voudrais revoir celui à gauche de la carte.

Le pronom démonstratif *celui* a comme « antécédent » l'icône affichée sur l'écran lors de l'activité S1 et non un antécédent présent dans le co-texte.

Ellipse : l'ellipse est un procédé linguistique qui consiste à supprimer une partie d'un énoncé sans en altérer le sens, sens qui peut être reconstruit par le récepteur.

Exemple 11
U1 : Quel est le tirant d'eau du Clémenceau ?
S1 : ....
U2 : *et son tonnage ?*

L'énoncé U2 est elliptique ; sa forme reconstituée serait après résolution du pronom possessif *son* : *Quel est le tonnage du Clémenceau ?*.

Pour certains, l'ellipse est la forme ultime de l'anaphore car reconstituer l'énoncé complet revient souvent à rechercher dans le co-texte un texte qui sert d'antécédent.

2.2 Résolution des expressions référentielles : difficultés du traitement

Dans des conditions de communication normales, l'énonciation (ou sa rédaction) d'une expression référentielle a pour objectif que l'auditeur (le lecteur) puisse identifier avec succès le référent[6]. Cette relation expression référentielle-référent ainsi que les conditions de succès

[6]Nous laissons de côté les situations de communication telles les dialogues dans certaines pièces de théâtre où les auteurs jouent sur les ambiguïtés possibles entre expressions référentielles et référents.

de son identification a préoccupé et préoccupe encore philosophes, linguistes et informaticiens.

En ce qui concerne les descriptions définies, B. Russel propose une interprétation fondée sur la logique [1]. Une description définie singulière est notée par le iota inversé (codé ι ici) ; l'expression *le camping vert* se note :

$(\iota x camping(x) \wedge vert(x))$

Cette notation est alors définie dans la logique classique avec égalité de la manière suivante :

$$\phi(\iota x \varphi(x)) \equiv (\exists x \varphi(x) \wedge (\forall y \varphi(y) \Rightarrow y = x) \wedge \phi(x))$$

L'énoncé *le camping vert est ouvert* peut se traduire par

$$ouvert(\iota camping(x) \wedge vert(x))$$

qui se développe en

$$(\exists x camping(x) \wedge vert(x) \wedge (\forall y camping(y) \wedge vert(y) \Rightarrow y = x) \wedge ouvert(x))$$

Cette façon de considérer une description définie dote l'énoncé de condition de vérité. L'énoncé *le camping vert est ouvert* est vrai si et seulement si il existe un et un seul camping vert et qu'il est libre. Il est faux dans l'un des trois cas suivants :

- s'il y a un et un seul camping vert et qu'il n'est pas ouvert ;
- s'il y a plusieurs campings verts,
- s'il n'y a aucun camping vert.

Plusieurs reproches ont été formulés ([17]) à l'encontre de cette approche. Le premier concerne l'absence de considération de contexte : la plupart des descriptions définies se rapportent à un contexte qui permet de délimiter la recherche du référent[7]. D'autre part, certains prétendent que s'il n'existe pas de référent pour la description définie, il n'est pas pertinent de se poser la question de la vérité de l'énoncé. Enfin un autre défaut concerne l'absence de distinction entre l'usage attributif et l'usage référentiel (cf. plus bas) dans les expressions définies.

Ces remarques renvoient d'une certaine manière à la notion d'usage que nous allons considérer dans les sections suivantes. Elles ont aussi conduit certains chercheurs (Searle, Clark & Marshall, Perrault & Cohen) à proposer une approche qui se place dans le cadre général de la théorie des actes de langage. Les conditions de succès d'une description définie (considérée comme une action) portent alors sur l'état des connaissances de l'auteur de la description définie.

Les expressions référentielles qui utilisent des indices linguistiques tels que les pronoms suscitent également des débats notamment sur le rôle et la représentation des informations contextuelles.

[7] On retrouve ici un partie du débat sur l'existence de la pragmatique.

La résolution automatique des différents types d'ER est au cœur de nos préoccupations. Nous nous trouvons devant les mêmes interrogations que celles signalées ci-dessus avec en plus celles inhérentes au contexte informatique. Les incertitudes dues à la reconnaissance de la parole (homophones hétérographes, mots courts, ...), les ambiguïtés syntaxiques si faciles à lever pour un humain mais difficiles à détecter et à traiter par un programme, la performance des utilisateurs dans leurs activités de désignation sont autant d'éléments qui compliquent la tâche ; ils incitent aussi à utiliser les propositions et résultats déjà présents mais à le faire avec précaution.

Avant d'exposer les méthodes les plus connues, nous présentons quelques situations d'usage d'ER qui sont susceptibles de poser des problèmes dans le cadre d'une résolution automatique. Ces situations concernent essentiellement les ER description définie et les ER avec démonstratifs. Les éléments descriptifs qui suivent sont extraits en grande partie de le thèse de Manuélian [29] qui s'est intéressée à la génération des descriptions définies et qui dans ce cadre a étudié les différents usages des ER.

2.3 Précisions sur statut et fonction

2.3.1 Référence actuelle et référence virtuelle

Millner distingue deux types de référence pour le syntagme nominal : la référence virtuelle et la référence actuelle selon la cible référent visée. Pour pouvoir identifier le référent d'un groupe nominal, il est nécessaire de faire appel à ces deux types de référence. Dans un premier temps (aspect cognitif), on utilise la référence virtuelle du syntagme qui est proche du sens lexical du syntagme nominal sans son déterminant. On convoque ainsi les conditions et propriétés que doit remplir le référent pour être désigné par la description utilisée. Ainsi quand on parle de *bois*, on doit passer par la référence virtuelle de bois qui font qu'un bois est un bois (un bois est constitué d'arbres, a une surface, une lisière, ...). La référence actuelle est ce qui permet d'attribuer la séquence linguistique à un référent du monde extra linguistique. Elle n'a d'existence et de pertinence que dans un contexte d'interprétation et si le nom est dans la portée d'un déterminant. Dans notre cadre applicatif, la notion de référence virtuelle doit être prise en compte par l'intermédiaire du lexique et des contraintes sémantiques.

2.3.2 Descriptions attributives et référentielles

Donnellan distingue aussi deux types de descriptions définies : attributives et référentielles.

Description attributive : les descriptions attributives possèdent les caractéristiques suivantes :

– elles servent à dire quelque chose à propos d'un objet, quel qu'il soit, à partir du moment où il correspond à la description ;

– il y a une présupposition générale que quelque chose satisfait la description : si rien ne satisfait la description, la question *qui est le x qui y ?* n'a pas de réponse ;
– elles dénotent mais ne réfèrent pas, c'est-à-dire qu'elles ont un sens, une référence virtuelle, mais pas de référence actuelle.

Description référentielle : les descriptions référentielles possèdent des caractéristiques parallèles à celles énumérées pour les descriptions attributives :
– elles servent à ce que l'auditeur puisse identifier le référent dont le locuteur parle ;
– il y a une présupposition qui affirme qu'un objet particulier correspond à la description , la question q*ui est le x qui y ?* a une réponse sous forme de description définie ;
– elles dénotent et elles réfèrent : ces expressions ont à la fois une référence actuelle et une référence virtuelle.

D'après Manuélian il est impossible d'utiliser une description démonstrative de façon atributive. Il n'est donc pas nécessaire dans notre cadre de s'attaquer aux descriptions attributives ; on pourrait donc utiliser cette remarque pour rejeter ou conforter des résultats de reconnaissances (la présence du démonstratif impliquant une utilisation référentielle).

2.4 Usage et problèmes sous-jacents

Une analyse de corpus fait apparaître assez rapidement que dans certains cas il est difficile de déterminer antécédent et/ou référent pour des descriptions définies introduites par des articles définis ou des démonstratifs. Ceci est notamment dû d'une part à la manière dont les expressions référentielles apparaissent (première mention ou en reprise anaphorique) et d'autre part aux entités lexicales utilisées. Dans cette section nous recensons ces différents usages et les problèmes de traitement sous-jacents.

2.4.1 Emplois en première mention

Défini en première mention.
Nous reprenons ici les deux types d'utilisation en première mention proposée par Vieira [40] pour le défini : les utilisations situationnelles et les utilisations non-familières.

- Utilisations situationnelles : elles sont de deux types ; elles se rapportent soit au contexte immédiat du dialogue (co-texte, contexte sémantique), soit à l'ensemble des contextes globaux. Pour les premières, il faut que le référent visé soit physiquement présent dans le contexte. Ainsi, dans Géoral, on pourra employer le syntagme nominal défini *la mer* pour désigner la mer représentée sur l'écran. Ces utilisations sont appelées de façons diverses dans la littérature : *utilisation en situation visible*, *utilisation en situation immédiate*, *utilisation pragmatique*, *utilisation évoquée situationnellement*.
Pour les secondes, les expressions qui se rapportent au contexte général de la communication, comme *la météo* pendant une recherche d'un site touristique, le référent n'est pas forcément visible, ou présent physiquement dans la situation de communication, mais sont

des objets connus, identifiables uniquement par les deux locuteurs au moment où se passent les échanges. Ces utilisations sont appelées *utilisation en situation plus large* ou *utilisation inférée situationnellement*.

- Utilisations non familières. Elles sont aussi appelées *containing inferrable* par Prince. Il s'agit de descriptions définies permettant d'identifier uniquement le référent parce qu'elles en donnent une description complète, ou en lien avec un autre référent connu, comme par exemple dans Géoral : *le bord de ce bois*, *l'origine de la rivière*. L'interprétation de ce type de référence nécessitera des connaissances supplémentaires.

Démonstratif en première mention
Le démonstratif peut avoir deux types d'utilisation en première mention : la première est appelée exophore a-mémorielle, la seconde est l'exophore mémorielle.

-Exophore a-mémorielle (ou deixis in praesentia) : il s'agit de cas identiques aux utilisations situationnelles (situation visible ou immédiate) pour les définis : c'est-à-dire que le référent est présent physiquement dans la situation de communication. Généralement, ces emplois du démonstratif sont accompagnés d'un geste. Ainsi, dans Géoral, on peut imaginer une situation où l'usager commence une communication en désignant la représentation d'une route et en prononçant :
je voudrais les campings le long de ***cette*** *route*.

-Exophore mémorielle (ou deixis in absentia). Cette seconde utilisation du démonstratif en première mention est probablement plus rare ; l'expression référentielle renvoie à un objet présent uniquement en mémoire du locuteur et présuppose la connaissance de l'objet par son interlocuteur. Dans l'exemple suivant :
Je voudrais l'adresse de ***cette*** *fameuse abbaye*, le locuteur fait peut-être allusion à un échange se déroulant en parallèle ou à un fait dont on a parlé dans les média.

2.4.2 Utilisations coréférentielles directes

Définis en reprise directe
Selon Manuélian, la littérature anglo-saxonne ne distingue pas systématiquement les emplois coréférentiels directs (avec la même tête nominale dans l'antécédent et l'anaphore) des emplois coréférentiels indirects (avec une tête nominale différente dans l'antécédent et l'anaphore). Elle choisit de les distinguer afin de cerner plus précisément le phénomène de reprise par un syntagme nominal défini. Des auteurs (Fraurud, Hawkins, Prince) parlent respectivement d'emplois en mention subséquente, d'emplois anaphoriques ou de référents évoqués textuellement. D'autres auteurs (Clark, Sidner et Strand) différencient les reprises directes des autres et parlent de relations d'identité, de co-spécification ou de co-référence avec une tête identique. La littérature française, toujours selon Manuélian, distingue trois types de reprise directes (ou fidèles) :

- les reprises totalement fidèles : l'antécédent et la reprise contiennent le même nom et les mêmes modifieurs ;
- les reprises directes par un défini nu : la reprise ne comporte pas de modifieurs et utilise le même nom tête que la première mention ;
- les reprises par un défini modifié : la reprise et la première mention ont la même tête nominale, mais les modifieurs varient d'une mention à l'autre.

-Reprise totalement fidèle
Cette notion est identique à celle de reprise directe, le nom de tête du syntagme doit être le même que dans l'antécédent. Il semble que la justification des cas de reprise totalement fidèle soit en partie informationnelle : si la propriété dénotée par les modifieurs a une valeur dans l'argumentation, ou est importante à souligner dans la prédication où apparaît la seconde mention du référent, la reprise totalement fidèle est possible.

-Reprise par un défini nu
Pour Corblin, le défini est indépendant de la notion de reprise. En effet un défini n'est pas systématiquement interprété comme un syntagme anaphorique. Pourtant la seconde mention d'un objet par une description définie sans modifieurs peut être interprétée comme une reprise.

-Reprise par un défini modifié
Si les définis nus peuvent être interprétés comme des reprises d'une mention antérieure, Corblin montre qu'il est difficile de faire une telle interprétation si on ajoute des modifieurs dans la seconde mention car il est difficile d'établir une relation de coréférence entre les deux syntagmes.

La reprise par un défini nu est attesté en français mais n'est pas possible dans tous les contextes si l'antécédent est indéfini. Par ailleurs, la reprise par un défini modifié et dont les modifieurs sont différents de ceux de l'antécédent paraît impossible.

Démonstratifs en reprise directe.
Pour Corblin, le statut de reprise d'une mention est définitoire du démonstratif. Kleiber affirme que l'adjectif démonstratif se révèle être un vrai connecteur anaphorique. Pour Kleiber, la reprise fidèle démonstrative a un statut déictique. Le démonstratif permet de référer à un objet présent dans le contexte linguistique, et peut être équivalent à un déictique désignant un objet du contexte extra-linguistique.

Synthèse
La reprise directe est trouvée avec les deux déterminants. Le problème est donc de savoir dans quelles conditions apparaissent les deux déterminants. Malheureusement, selon Manuèlian, ils n'apparaissent pas exactement en distribution complémentaire. Les critères fournis par Kleiber et Corbin sont les suivants :

- le défini apparaît lorsqu'on peut établir un contraste avec un autre référent du contexte ;
- le démonstratif est utilisable à chaque fois sauf dans le cas où le référent est explicitement opposable à un autre de par sa catégorie ;
- le démonstratif est meilleur que le défini dans les cas où le prédicat associé à la reprise anaphorique constitue une rupture dans la continuité événementielle.

Dans notre contexte, la présence du geste (appui du démonstratif) ainsi que la notion de saillance visuelle apporteront des indices pour déterminer la situation d'usage.

2.4.3 Utilisations coréférentielles indirectes

Reprise définie indirecte.
La reprise définie infidèle est assez peu mentionnée dans la littérature française. Pour Corblin, il semble que seuls quelques noms de qualité puissent être employés en reprise indirecte avec le défini (i.e, *la crapule*, *l'imbécile*, *le jeune homme*, ...). La reprise indirecte du référent par une description définie donne alors lieu à deux types d'interprétation :

- la reclassification est interprétée comme occasionnelle si la reprise occupe la fonction de sujet grammatical dans la phrase ;
- la reclassification est interprétée comme permanente si la reprise occupe une autre fonction grammaticale que la fonction sujet.

Dans la littérature anglo-saxonne, on trouve de nombreuses mentions de la reprise définie indirecte. Pour Clark, une reprise définie indirecte peut être une *pronominalisation*, c'est-à-dire un usage où le nom employé dans la reprise est plus générique que son antécédent (*une rivière ... un cours d'eau*), ou un usage où la reprise donnera le type d'un objet mentionné en premier lieu par un nom propre (*Beg Leguer ... la plage*) ; il s'agira d'un épithète quand les noms n'auront pas de relation sémantique (*un homme ... le salaud*). Pour Strand, il s'agira respectivement de cas de *généralisation* quand la reprise est un hyperonyme de l'antécédent et de *redescription* si l'antécédent est un nom propre ou s'il n'y a pas de relation sémantique entre les deux noms. Strand ajoute à ces emplois des spécifications quand la reprise est un hyponyme de l'antécédent (*la voiture ... la berline*) et des élargissements dans un cas comme : *un homme et une femme ... le couple*.
Les reprises définies indirectes apparaissent dans le même type de contexte que celui des reprises directes. Un moyen de les classer est d'utiliser la relation lexicale qui unit le nom contenu dans le syntagme de reprise et le nom servant d'antécédent. Même s'il existe des cas où il n'y a pas de relation lexicale connue entre l'antécédent et l'anaphore, il semble important (contrairement au démonstratif) de pouvoir établir un lien entre les deux syntagmes. Ce lien passe soit par le lexique, soit par des connaissances du monde dans le cas de redescription et des épithètes par exemple.

Reprise démonstrative indirecte
Pour Corblin, la référence virtuelle du syntagme n'est pas prise en compte dans l'interprétation d'un syntagme démonstratif. Ceci permet au locuteur de re nommer un antécédent pratiquement à volonté à partir du moment où l'interlocuteur est capable de reconstituer le lien entre l'antécédent et l'anaphore. Les exemples fournis par Corblin concernent la mé-

taphore et l'attribution de propriétés (sans spécifier que l'attribution de propriétés doit constituer ou non un apport d'information). D'autre part, Manuèlian note que quand le lien de coréférence n'est pas purement lexical et qu'il y a plusieurs antécédents possibles, il peut être difficile de choisir entre les antécédents concurrents ; des paramètres de saillance ou de récence peuvent être nécessaires.
Enfin, Manuèlian note également des emplois appelés *simplification de l'antécédent* (par Wiederspiel) pour qui la tendance générale de l'anaphore est à la simplification de l'antécédent. La simplification peut être conceptuelle, si le terme employé dans la reprise est plus générique que celui employé dans l'antécédent ou formelle, si l'antécédent est un groupe verbal (on est dans un cas de réification), une proposition et la reprise un groupe nominal. Ce type d'emploi ne sera pas pris en compte dans notre cadre.

2.4.4 Anaphore associative

L'anaphore associative est un phénomène où l'anaphore et l'antécédent ne sont pas coréférents.

Exemple 12
Au guichet d'une poste :
S1 : Allez à la cabine 8.
U1 : *Le téléphone* est cassé

Dans cet exemple, *le téléphone* est associé spontanément à celui présent dans la cabine (téléphonique) 8. Sans la mention explicite de la cabine, on ne pourrait pas utiliser le défini et identifier le référent. L'anaphore associative est donc un mécanisme qui permet d'introduire un nouveau référent en s'appuyant sur un antécédent. Ce référent nouveau doit entretenir une relation précise avec son antécédent, relation qui nécessite pour la découvrir des inférences et des connaissances souvent encyclopédiques de la part de l'interlocuteur. Plusieurs typologies des relations ont été proposées ; nous retenons celle de Kleiber.

Anaphore associative entre membre et collection. Il s'agit de la relation entre un élément et un ensemble comme dans :

Exemple 13
J'ai rencontré un couple hier. L'homme était réellement stupide.

Dans cet exemple on interprète le GN *l'homme* comme référant à l'homme faisant partie du couple parce que nos connaissances du monde nous disent que l'on désigne généralement par le mot *couple* un ensemble composé d'un homme et d'une femme.

Anaphore méronymique. La relation méronymique est la relation *partie-tout*. L'anaphore est une partie de l'antécédent, la relation est illustrée par le couple *cabine-téléphone* comme dans l'exemple 12.

Anaphore locative. Le référent de l'anaphore est « situé » dans le référent de l'antécédent et la relation est illustrée par la relation entre *village* - *l'église* de l'exemple classique :

Exemple 14
Nous arrivâmes dans un village. **L'église** était située sur une hauteur.

Anaphore fonctionnelle.
Cette relation s'établit quand le nom anaphorique présente un trait fonctionnel avec l'antécédent, comme dans :

Exemple 15
Le livre se vend se vend bien. **L'auteur** est très satisfait.

Anaphore actancielle. L'anaphore est un actant de l'événement décrit précédemment :

Exemple 16
Jean a été assassiné hier. Le meurtrier court toujours.

Dans notre contexte, les anaphores méronymique et locative sont les plus probables.

2.4.5 Autres phénomènes

Anaphore zéro
Ce phénomène est lié à l'ellipse. L'ellipse pouvant survenir sur toutes parties de phrases, nous allons trouver plusieurs types d'anaphores zéro :

-Anaphore pronominale zéro
Un pronom anaphorique est omis mais est néanmoins compris.

Exemple 17
Le pronom anaphorique ϕ sujet est omis mais est néanmoins compris.
Fred tira et ϕ marqua le but.

-Anaphore nominale zéro
Elle survient quand seul le nom de tête, et non tout le groupe nominal, est omis ; le lien référenciel est établi par les modifieurs non omis.

Exemple 18
Georges a acheté une grosse boîte de chocolats et peu ϕ ont survécu en fin de journée.

Cataphore
La cataphore survient quand une référence est faite sur une entité mentionnée plus loin dans le co-texte.

Exemple 19
Elle est maintenant célébre comme son ex-boyfriend. Du désert du Kazakhstan aux mers du Sud de Tonga, tout le monde connaît *Monica Levinsky.* ([31]

Mitkov note que les références cataphoriques sont typiques des genres littéraire et journalistique. Dans un premier temps nous ne traiterons pas ce type de phénomène compte tenu du type de texte traité et de la difficulté de sa prise en compte automatique.

Fausse anaphore

-Pronom pléonastique
Le pronom *il* peut être utilisé de manière non anaphorique, usage impersonnel dit aussi pléonastique ([27]).

Exemple 20
Il pleut.
Il semble nécessaire de faire du sport régulièrement.

Ces types d'emploi peuvent être répertoriés (formes figées, utilisation de verbes modaux) mais leur traitement en TALN ne semble pas si évidente [31].

3 Algorithmes de traitement des anaphores pronominales

3.1 Présentation

La réalisation des premiers systèmes informatiques, dans les années soixantes dix, avec comme application l'interrogation de bases de données s'est trouvée, très rapidement confrontée au problème de la résolution des anaphores pronominales, phénomène omni-présent dans un tel contexte. Les premiers algorithmes mis au point (STUDENT [3], SHRDLU [42], [41], [20]) ont plutôt utilisé des règles heuristiques fondées sur l'observation de situations courantes. Très rapidement, l'apport de connaissances supplémentaires (principalement syntaxiques mais pas seulement) est apparu comme intéressant. Hobbs a ainsi proposé une approche dite « naïve » qui utilise les résultats d'une analyse syntaxique simple et un algorithme de parcours d'arbre. Dans la même lignée mais avec des observations syntaxiques plus fines sur des corpus, Lappin et Leass [27] ont proposé une méthode fondée sur la structure syntaxique des énoncés, une mesure de saillance et un modèle attentionnel. Ces différents travaux préconisent également l'usage de connaissances sémantiques pour tenter de résoudre les situations ambiguës. Cependant, cette préconisation n'est pas toujours complètement formalisée.
En opposition avec le « tout syntaxique », des approches sémantiques ont été introduites fondées sur la DRT[8] [25] ou sur les graphes conceptuels. La prise en compte des caractéristiques intentionnelle et de cohérence des textes et dialogues ont également amené Grosz et

[8]Discourse Representation Theory.

Sidner à suggérer une théorie, la théorie du centre [15], pour permettre de suivre les déplacements du foyer d'attention (le centre) dans un texte ou un dialogue et ainsi résoudre plus facilement les anaphores pronominales. Cette théorie fait toujours l'objet de développements.

Nous allons examiner quelques-uns de ces algorithmes et propositions. Notre propos est de mettre en évidence les éléments qui nous semblent réutilisables, aucun de ces algorithmes ou propositions n'apparaissant comme suffisamment complet ou général pour pouvoir être utilisé seul et directement.

3.2 Approche de Hobbs (1978)

Hobbs [24] propose une approche syntaxique, dite naïve[9], pour la résolution des pronoms $3^{\text{ième}}$ personne dans des textes écrits en anglais. L'algorithme opère sur les arbres d'analyse syntaxique qui représentent les structures grammaticales correctes des énoncés. La recherche d'un antécédent est effectuée dans l'arbre syntaxique de l'énoncé courant de gauche à droite avec un parcours en largeur d'abord et donne la préférence aux antécédents les plus proches du pronom. Lorsqu'aucun antécédent n'est trouvé dans l'arbre de la phrase courante, on parcourt les arbres syntaxiques des phrases précédentes, en commençant par la phrase immédiatement précédente, ce qui traduit la préférence pour des antécédents proches plutôt qu'éloignés.

Pour illustrer son algorithme Hobbs utilise la grammaire hors contexte suivante (adaptée pour une partie du français) qui permet de générer les structures des arbres d'analyse syntaxique :

(1) $S \rightarrow NP\ VP$
(2) $NP \rightarrow (DET)N\ (PP|REL)^* \mid pronom$
(3) $DET \rightarrow article \mid possessif$
(4) $N \rightarrow nom(PP)^*$
(5) $PP \rightarrow preposition\ NP$
(6) $REL \rightarrow adverbe \mid relatif\ S$
(7) $VP \rightarrow (NP)verbe(NP)(PP)^*$

Un terminal en italique renvoie aux mots lexicaux de la catégorie exprimée par la dénomination du terminal.
Hobbs spécifie son algorithme comme suit :

1. Commencer au nœud NP qui domine immédiatement le pronom dans l'arbre d'analyse de la phrase S.
2. Remonter dans l'arbre jusqu'au premier nœud NP ou S rencontré. Appeler ce nœud X et le chemin qui y mène P.

[9]Elle est qualifiée de naïve en raison de sa simplicité car elle n'utilise qu'un algorithme de parcours d'arbre.

3. Parcourir toutes les branches sous le nœud X à la gauche du chemin P de gauche à droite en largeur d'abord. Proposer comme antécédent tout nœud NP rencontré qui a un nœud S ou NP entre lui et X.
4. Si le nœud X est le plus haut nœud S dans la phrase, parcourir les arbres des énoncés précédents dans le co-texte dans l'ordre de recense, le plus récent d'abord ; chaque arbre est examiné de gauche à droite en largeur d'abord, et quand un nœud NP est rencontré, il est proposé comme antécédent. Si X n'est pas le nœud le plus haut dans la phrase, aller à l'étape 5.
5. Du nœud X, remonter dans l'arbre jusqu'au premier nœud S ou NP rencontré. Appeler ce nœud X et le chemin qui y mène P.
6. Si X est un nœud NP et si le chemin P ne passe pas par le nœud N que X domine immédiatement, proposer X comme antécédent.
7. Parcourir toutes les branches sous le nœud X à gauche du chemin P de gauche à droite en largeur d'abord. Proposer tout nœud NP rencontré comme antécédent.
8. Si X est un nœud S, parcourir toutes les branches du noeud X à la droite du chemin P de gauche à droite en largeur d'abord, mais ne pas aller sous les nœuds NP ou S rencontrés. Proposer tout nœud NP rencontré comme antécédent.
9. Aller à l'étape 4.

Dans son algorithme, quand Hobbs parle de proposer un nœud comme antécédent, la proposition peut être refusée sur la base de l'accord en genre et en nombre entre pronom et antécédent. Hobbs spécifie qu'il peut être fait appel à d'autres contraintes de sélection à partir des deux principes suivants :

– **principe A** : un pronom non-réflexif ne peut pas trouver son antécédent dans la même phrase simple. Cela prédit par exemple l'impossibilité de coréférence entre *Kheops* et *le* dans 22 par rapport à la coréférence entre *Kheops* et *se* dans 21.

Exemple 21
Il y a Ramses qui n'est encore pas couché.
Kheops se sent un peu à lŠétroit en ce moment.

Exemple 22
Il y a Ramses qui n'est encore pas couché.
Kheops le sent un peu à l'étroit en ce moment.

– **principe B** : l'antécédent d'un pronom (non réfléchi et non réflexif) doit précéder ou c-commander le pronom. La notion de c-commande est définie comme suit : un nœud N1 commande un nœud N2 si N1 ne domine pas N2 et N2 ne domine pas N1 et si le premier nœud branchant dominant N1 domine aussi N2. Ce principe prédit par exemple la possibilité de coréférence en 23 et lŠimpossibilité de coréférence en 24.

Exemple 23
Après qu'il ait dévalisé la banque, Jean quitta la ville.

Exemple 24
Jean le lave.

Dans le déroulement de l'algorithme, un refus de proposition fait passer à la suite du parcours ou à l'étape suivante. L'algorithme a été évalué sur 300 pronoms non déictiques, issus de trois types de textes différents (technique, journalistique, littéraire). Les pronoms traités étaient *he*, *she*, *it* et *they*. Hobbs a étudié la distribution des pronoms et ses antécédents dans les textes et a défini les ensembles candidats C_0, C_1 ... C_N avec $C_0 \subset C_1$, $C_1 \subset C_2$, ...

C_0 = (a) l'ensemble des entités de la phrase en cours et de la phrase précédente si le pronom apparaît avant le verbe principal, ou (b) l'ensemble des entités seulement de la phrase en cours si le pronom apparaît après le verbe principal.
C_1 = l'ensemble des entités de la phrase en cours et de la phrase précédente.
C_N = l'ensemble des entités de la phrase en cours et des N phrases précédentes

Hobbs observe que 90% de tous les antécédents sont dans C_0 et que 98% sont dans C_1. Cela n'empêche pas l'existence d'antécédents beaucoup plus éloignés, ainsi dans l'un des exemples du corpus de test, un antécédent était neuf phrases plus loin que le pronom. Il a aussi observé que le pronom *it*, spécialement dans des textes techniques, peut avoir un nombre d'antécédents relativement élevé dans une même phrase, dans l'un de ces cas *it* avait 13 antécédents. Hobbs a ainsi constaté que même en limitant le domaine de recherche des antécédents aux N phrases précédentes, on peut trouver des cas avec des dizaines d'antécédents possibles [23]. Le taux de réussite pour la résolution est de 88.3%. La version de cet algorithme qui utilise des contraintes sélectives (mentionnées ci-dessus) fournit un taux de réussite de 91.7%. Hobbs juge que ces taux de reussite sont quelque peu trompeurs parce que dans plus de la moitié des cas il n'y avait qu'un seul antécédent plausible [31]. Pour cette raison l'algorithme a été testé sur des exemples dans lesquels il y avait au moins deux antécédents plausibles pour chaque pronom. 81.8% des conflits sont résolus par l'usage combiné de l'algorithme et des contraintes de sélection, ce qui est un taux intéressant. Hobbs en conclut que l'approche naïve est très bonne.

Cet algorithme ainsi que les observations qui ont conduit à le construire servent toujours de référence (voir également l'ouvrage de J. Allen [1]). Cependant, il faut noter qu'il est très difficile d'extrapoler ces bons résultats dans notre cas. Le fait de travailler sur du texte écrit et non sur des échanges dialogiques, les hypothèses très fortes sur les arbres syntaxiques (complets et sans erreur) font que le taux de succès pourrait être beaucoup plus bas.

3.3 Approche de Lappin et Leass (1994)

Lappin et Leass proposent un algorithme pour trouver les antécédents des pronoms personnels ($3^{\text{ième}}$ personne). L'algorithme travaille sur des représentations syntaxiques et utilise

des mesures de saillance déduites de structures syntaxiques générales ainsi qu'un modèle de dialogue attentionnel simple et dynamique pour gérer le choix d'un antécédent à partir d'une liste de GN candidats. Aucune connaissance sémantique ou encyclopédique n'est utilisée. Cet algorithme, dénommé RAP[10], contient les composantes principales suivantes :

1. Un filtre syntaxique intrasentenciel pour éliminer les coréférences entre pronom et GN sur des fondements syntaxiques.
2. Un filtre morphologique pour éliminer les coréférences ente pronom et GN pour des raisons de non-accord sur personne, genre ou nombre.
3. Une procédure pour identifier les pronoms pléonastiques (sémantiquement vide).
4. Un algorithme pour identifier l'antécédent possible d'un pronom réflexif ou réciproque dans la même phrase (comme dans *Ces deux élèves* ***s'****aident toujours* ou dans *Nous* ***nous*** *parlons*).
5. Une procédure pour assigner des valeurs aux différents paramètres de saillan-ce d'un GN. Elle assigne des poids élevés aux GN sujets par rapports aux GN non sujets, aux compléments d'objets directs par rapport aux autres compléments, etc (cf. Tableau 1).
6. Une procédure pour identifier les GNs anaphoriquement liés comme une classe d'équivalence pour laquelle une valeur globale de saillance est calculée comme étant la somme des valeurs de saillance de ses éléments.
7. Une procédure de décision pour sélectionner l'élément pertinent de la liste des antécédents candidats pour un pronom.

TAB. 1 – Scores de saillance en fonction de critères syntaxiques [27]

Type du Facteur	**Seuil initial**
Récence phrastique	100
Emphase sur le sujet	80
Emphase existentielle	70
Emphase accusative	50
Objet indirect	40
Tête dŠun GN	80
Emphase non adverbiale	50

LŠalgorithme proprement dit est le suivant :

1. créer une liste pour tous les GN de la phrase courante ; les GN sont classés selon leur type (indéfini, défini, pronom pléonastique, pronom personnel, pronom réflexif, indéfiniẼ)

[10]Resolution of Anaphora Procedure.

2. **pour chaque** GN de la phrase courante :

   (a) **si** GN = indéfini | défini **alors**
   calcul de la saillance

   (b) **sinon si** GN = pronom réflexif **alors**
   calculer une liste de paires GN - pronom pour lesquelles il peut y avoir coréférence : si plusieurs possibilités, l'antécédent est choisi sur le critère de saillance des GNs
   sinon
   si GN = pronom personnel de 3$^{\text{ième}}$ personne **alors**

      i. calculer une liste de paires GN - pronom pour lesquelles il ne peut pas y avoir co-référence

      ii. créer la liste des antécédents possibles : celle-ci contient le référent discursif le plus récent

      iii. modifications locales de la saillance : si le GN antécédent se trouve après le pronom, la saillance décroît (pénalisation des cataphores). Si le GN antécédent a le même rôle grammatical que le pronom, sa saillance augmente

      iv. définir un seuil de saillance et filtrage

      v. appliquer le filtre morphologique
      si plusieurs candidats **alors**
      choix sur la saillance du GN antécédent
      si égalité de saillance **alors**
      choix selon proximité
      fsi
      fsi

   fsi
   fsi
   fsi

fin pour chaque

Les principales caractéristiques de cet algorithme sont l'utilisation de bons filtres syntaxiques et morphologiques (plus fins que ceux de J. Hobbs), une prise en compte des pronoms pléonastiques, une mesure de la saillance fondée sur d'autres critères syntaxiques que la seule linéarité, le choix sur des critères de proximité en cas d'ambiguïté sur la saillance et la préférence des anaphores intraphrastiques sur les anaphores interphrastiques. Tout comme l'algorithme de J. Hobbs, RAP n'exige pas des connaissances sémantiques ou pragmatiques. Il faut également noter que les poids de saillance ont été déterminés par examen de corpus. Ils dépendent donc de la langue et du genre de texte.
Différentes évaluations de l'algorithme ont été menées. Une évaluation sur 354 phrases extraites d'un texte technique a donné un taux de succès de 86% pour un total de 360 pronoms.

Ce taux dépasse donc de 4% celui obtenu par l'algorithme de J. Hobbs sur les mêmes données (cf. Tableau 2). Cette performance est essentiellement due à un meilleur traitement des anaphores intraphrastiques qui constituent, selon les auteurs, 80% des pronoms du corpus. Ceci indique que la prépondérance fournit une base plus fiable pour le choix de l'antécédent que la procédure de recherche de Hobbs dans le domaine des textes où les deux algorithmes ont été testés.

TAB. 2 – Comparaison des algorithmes de Hobbs (1978) et de Lappin et Leass (1994)

	Total	**Interphrastique**	**Intraphrastique**
Occurrences pronoms	360	70	290
Cas corrects (RAP)	310 (86%)	52 (74%)	258 (89%)
Cas corrects (Hobbs)	295 (82%)	61 (87%)	234 (81%)

3.4 Approche de Mitkov

Cet algorithme [31] n'exige pas une analyse syntaxique complète ; il s'applique sur un texte déjà soumis à une analyse morphologique. Il utilise une analyse syntaxique de surface (shallow parsing) adaptée pour l'extraction des GN. Les GN considérés appartiennent à la phrase courante ainsi qu'aux deux phrases précédentes. Les GN traités sont simples et ne contiennent pas de phrases complexes imbriquées.

Afin de trouver l'antécédent d'un pronom, l'algorithme établit la liste des GN de la phrase courante et des deux phrases précédentes, exclut ceux qui ne s'accordent pas en nombre ou en genre. Ensuite, on applique les indicateurs d'antécédence aux NP qui ont passé le filtre genre et nombre. Ces indicateurs agissent soit comme amplificateur soit comme impédance. Les indicateurs amplificateurs attribuent un score positif à un GN traduissant ainsi une probabilité positive que le GN puisse être l'antécédent du pronom. A contrario, les indicateurs d'impédance attribuent un score négatif à un GN traduisant ainsi un manque de confiance que le GN soit l'antécédent du GN. La plupart des indicateurs sont indépendant du genre de texte et sont liés aux phénomènes de cohérence (tels que saillance et distance) ou aux structures, tandis que certains dépendent du genre de texte.

Les indicateurs amplificateurs sont les suivants :
- *Premier GN* : un score de +1 est assigné au premier GN de la phrase.
- *Verbes d'indication* : un score de +1 est assigné aux GN qui suivent immédiatement certains verbes anglais d'indication appartenant à un ensemble prédéfini comme : analyse, assess, check, consider, cover, define, describe, develop, discuss, examine, explore, etc.
- *Répétition lexicale* : un score de +2 est assigné aux GN répétés deux fois ou plus dans le paragraphe où apparaît le pronom et +1 aux GN répétés une fois dans ce paragraphe.

- *Préférence de titre* : un score de +1 est assigné aux GN figurant dans le titre de la section.
- *Localisation identique* : un score de +2 est assigné aux GN qui ont une forme de localisation identique à celle du pronom, comme le GN *the key* dans l'exemple suivant (tiré de [31]) :

  Exemple 25
  Press *the key* down and turn the volume up ... Press *it* again.

- *Référence immédiate* : un score de +2 est assigné aux GN qui apparaissent dans des constructions de la forme :
  (You) V_1 NP ... *con* (you) V_2 it (*con* (you) V_3 it
  où *con* $\in$ {and, or, before, after, until}

- *Instructions en séquence* : un score de +2 est appliqué au GN dans la position de GN_1 des constructions de la forme :

  To V_1 GN_1, V_2 GN_2. (phrase). To V_3 it, V_4 GN_4

  où le GN_1 est l'antécédent plausible pour de l'anaphore *it* (GN_1 a un score de 2).
- *Préférence de terme* : un score de +1 est assigné aux GN identifiés comme termes du domaine du texte.

Les indicateurs d'impédance :
- *Non défini* : un score de -1 est assigné aux GN indéfinis
- *GN prépositionnel* : un score de -1 est assigné aux GN qui apparaissent dans des syntagmes prépositionnels.

Enfin un indicateur, *la distance référentielle*, peut pénaliser ou améliorer les chances d'un candidat d'être sélectionné comme l'antécédent d'un pronom en fonction de la distance (en terme de clauses ou de frontières) entre le pronom et l'antécédent. Les GN dans la clause précédente (mais même phrase) que celle du pronom ont un score de +2, dans la phrase précédente +1, dans la phrase encore antérieure 0, dans les autres phrases antérieures -1.

L'algorithme peut être décrit de manière informelle :

1. Examen de la phrase courante et des deux phrases précédentes (si elles sont disponibles). Examiner les GN qui sont à gauche de l'anaphore.
2. Sélectionner de la liste des GN identifiés seulement ceux qui s'accordent en genre et nombre avec l'anaphore et les grouper dans l'ensemble des candidats potentiels.
3. Appliquer les indicateurs d'antécédence à chaque candidat potentiel et assigner les scores ; proposer le candidat avec le plus haut score cumulé. Si deux candidats ont le même score, proposer le candidat qui a le plus haut score pour l'indicateur référence immédiate. Si cela ne suffit pas, proposer le candidat qui a le plus haut score pour l'indicateur localisation identique. Si cet indicateur fournit une égalité ou n'est pas

pertinent, sélectionner le candidat qui a le plus haut score pour l'indicateur verbes d'indication. Si enfin, cela ne donne rien, prendre le candidat le plus récent.

La stratégie de résolution peut être illustrée à partir de l'exemple suivant (tiré de [31]). Le but est de trouver l'antécédent de *it* dans ce texte :

> ***Positioning the original : Standard Sheet Original***
> *Raise the original cover. Place the original face down on the original glass so that it is centrally aligned against the original width scal. The center of the original must be aligned with the arrow marking on the original width scale.*

Les étape 1 et 2 de l'algorithme génèrent l'ensemble des candidats potentiels : {original, cover, original, original glass}

L'étape 3 assigne les scores suivants aux candidats

original cover

1 (premier NP de la phrase) + 0 (verbe d'indication) + 0 (répétition lexicale) + 0 (Préférence de titre) + 0 (Localisation) + 0 (référence immédiate) + 0 (instructions en séquence) + 1 (préférence de terme) + 0 (Non défini) + 0 (GN prépositionnel) + 1 (distance référentielle) = 3

original

1 (premier GN de la phrase) + 0 (verbe d'indication) + 1 (répétition lexicale) + 1 (Préférence de titre) + 0 (Localisation) + 0 (référence immédiate) + 0 (instructions en séquence) + 1 (préférence de terme) + 0 (Non défini) + 0 (GN prépositionnel) + 2 (distance référentielle) = 6

original glass

0 (premier NP de la phrase) + 0 (verbe d'indication) + 0 (répétition lexicale) + 0 (Préférence de titre) + 0 (Localisation) + 0 (référence immédiate) + 0 (instructions en séquence) + 1 (préférence de terme) + 0 (Non défini) + (-1) (GN prépositionnel) + 2 (distance référentielle) = 2

Donc le GN *the original* (score 6) est sélectionné comme antécédent de *it*.

Mitkov souligne qu'en dépit de la non utilisation de connaissances syntaxiques et sémantiques (hormis la liste de termes du domaine [31]), les résultats de l'algorithme sont comparables à ceux des méthodes utilisant des connaissances syntaxiques telles que celle de Lapin et Leass. Une évaluation de l'algorithme sur des textes techniques en anglais comportant au total 223 pronoms anaphoriques a donné 89.7% comme taux de réussite, ce qui

dépasse le score de RAP (86%). Mitkov a mené d'autres expérimentations et comparaisons qui l'amènent à penser beaucoup de bien de son approche.

Des remarques peuvent cependant être formulées quant au bien-fondé de plusieurs caractéristiques de la méthode. Ainsi Susanne Salmon-Alt dans sa thèse [35], met en lumière certaines interrogations sur l'approche de Mitkov : « Pourquoi un indéfini a-t-il moins de chances d'être l'antécédent d'un pronom ? D'où vient la liste des verbes ? Comment justifier la pondération des scores ? ». Elle poursuit « ce type d'approche comporte deux risques : d'un point de vue pratique, les systèmes implémentant une telle approche sont difficilement maintenables et peu évolutifs, car ils fournissent toujours une réponse, ne signalent jamais de problèmes et sont incapables de fournir des indications sur la source éventuelle des problèmes. D'un point de vue théorique, ils ne valident aucune hypothèse scientifique et ne contribuent en rien à l'explication des processus linguistiques et cognitifs sous-jacents à la résolution des anaphores ». Certaines de ces critiques sont valides (réglage du poids des indicateurs), d'autres ne sont pas vraiment fondées car elles ne prennent pas en compte les objectifs initiaux du travail. De notre point de vue, cet algorithme reste l'une des approches la plus puissante puisqu'il dépasse visiblement le RAP. Cependant, le réglage des paramètres est problématique et demanderait à faire l'objet d'études.

3.5 Théorie du Centrage (Centering)

Cette théorie est en grande partie issue des travaux de Grosz et Sidner [16] sur la modélisation du dialogue. Dans ces travaux, Grosz et Sidner proposaient un modèle de dialogue fondé sur trois structures : linguistique, attentionnelle et intentionnelle qui permettent d'assurer conjointement la cohérence entre les énoncés d'un dialogue. Cette idée de cohérence se trouve traduite dans la théorie du centrage pour permettre le rattachement des anaphores. La théorie du centrage [4, 16], qui concerne la résolution pronominale, est fondée sur l'idée que chaque énoncé met en avant une entité thématiquement plus importante appelée le **centre**. Ceci impose certaines contraintes dans la construction des énoncés sur l'usage des expressions référentielles et en particulier sur l'usage des pronoms. La théorie soutient que la cohérence d'un discours dépend dans une certaine mesure de la façon dont le choix des expressions référentielles se conforme aux propriétés du centrage. Cette théorie a suscité beaucoup de développements ultérieurs.

Dans la théorie du centrage, un discours (texte, dialogue) est constitué de segments. Un segment D est constitué d'une séquence d'énoncés S_1, S_2, .. S_N. Chaque énoncé S possède un ensemble des prochains centres potentiels le forward-looking Cf(S, D) qui correspond aux entités du discours évoqués dans l'énoncé. Chaque énoncé (sauf le premier) dans un segment possède un seul centre appelé backward-looking Cb(S). Le backward-looking Cb(S) est un élément du forward-looking Cf(S, D) et représente l'entité du discours dont l'énoncé parle. L'entité Cb relie l'énoncé courant au discours précédent : il focalise sur une entité qui a déjà été introduite. L'ensemble Cf est partiellement ordonné. L'ordonnancement se fait sur la base de critères grammaticaux, avec lesquels les entités en position sujet sont plus saillantes

que celles en position objet direct, elles-mêmes mises en avant par rapport aux objets indirects et aux circonstanciels. L'élément placé en tête du Cf(S) est appelé centre préféré. Le centre préféré dans la phase courante S_n (noté $Cp(U_n)$) est le centre backward-looking le plus probable de la phase suivante. Cette notion de centre, ne peut pas être définie syntaxiquement, c'est-à-dire que le syntaxe d'une phrase S ne peut pas déterminer quel groupe nominal réalise le Cb(S) et les Cf(S) [14]. Ce sont des facteurs syntaxiques, sémantiques et pragmatiques qui interviennent dans l'interprétation et la génération des phrases nominales définies dans le discours. L'exemple suivant montre l'intérêt de ces facteurs :

Exemple 26
Qui a vu Max hier ? (1)
Max a vu Rosa. (2)

Est-ce que quelqu'un a vu Rosa hier ? (3)
Max a vu Rosa. (4)

Bien que les énoncés (2) et (4) soient identiques, Cb(2) est Max et Cb(4) est Rosa.

Grosz définit pour le centre trois types de transition entre énoncés :
- continuation : le centre backward-looking de S_{i+1} est le même que celui de S_i (s'il existe) et constitue en même temps le centre préféré de S_{i+1}. Il constitue également le candidat le plus vraisemblable pour $Cb(S_{i+2})$.
- maintien : le centre backward-looking de S_{i+1} est le même que celui de S_i (s'il existe), mais il ne constitue plus en même temps le centre préféré de S_{i+1} (il n'appararaît plus en tête de liste) ;
- changement : le centre backward-looking de S_{i+1} n'est plus le même que celui de S_i.

Le pouvoir prédictif de la théorie du Centrage vient de la définition d'un ordre de préférence sur les transitions entre deux segments (*continuation* > *maintien* > *changement*) et d'une règle postulant que s'il y a pronominalisation en S_{i+1}, alors le centre backward-looking de S_{i+1} doit être pronominalisé [35]. L'application de ces principes doit produire des discours cohérents.

L'algorithme BFP.
L'algorithme « BFP » (pour Brennan, Friedman et Pollard), proposé par Brennan et al. [5] est une version algorithmique de la théorie du centrage qui a fait l'objet de différentes modifications, en particulier par M. Walker et al. en 1994. Cet algorithme reprend la distinction des différents types de transitions. En ce qui concerne la transition de type *changement*, Walker (1994) introduit une distinction supplémentaire entre *changement majeur* et *changement mineur*. Le Tableau 3 résume les transitions en fonction de l'évolution des centres entre deux phrases S_i et S_{i+1}. L'algorithme BFP redéfinit l'ordonnancement des transitions ainsi : *continuation* > *maintien* > *changement mineur* > *changement majeur*. La préférence est donc donnée à la stabilité thématique qui se traduit par une transition de type *conti-*

nuation.

TAB. 3 – Définition des transitions par l'algorithme BFP

	$Cb_i = Cb_{i+1}$ (ou pas de Cb_i)	$Cb_i \neq Cb_{i+1}$
$Cb_{i+1} = Cp_{i+1}$	continuation	changement mineur
$Cb_{i+1} \neq Cp_{i+1}$	rétention	changement majeur

L'algorithme proprement dit est le suivant :

1. Pour deux phrases données, calculer toutes les combinaisons Cb_i/Cf_{i+1} qui respectent l'accord en nombre et genre
2. Filtrer les combinaisons obtenues
   (a) par des contraintes sur les anaphores intraphrastiques (cf. les deux principes issus de la grammaire générative et utilisés par l'algorithme de Hobbs)
   (b) par des contraintes sur les restrictions de sélection (compatibilité sémantique)
   (c) par la règle de centrage sur la pronominalisation : Si un élément de la liste des centres forward-looking de la phrase S_i est pronominalisé dans la phrase S_{i+1}, alors le centre en arrière de S_{i+1} doit être pronominalisé
3. Classer les solutions restantes selon les types de transition
4. Choisir selon les préférences sur les types de transition

L'apport principal de la théorie du Centrage est sa simplicité calculatoire, permettant des implémentations relativement rapides. De plus, la caractérisation des relations entre les énoncés du discours apporte des indices intéressants dans le cadre de la structuration du discours, en particulier sur le plan de la délimitation d'espaces textuels cohérents. Il s'agit, en revanche, d'une approche extrêmement locale, puisque l'on reconstruit un nouveau contexte sous forme de liste de centres pour chaque phrase et ce contexte est limité à un seul segment. C'est une approche purement linéaire dans ces relations entre énoncés, négligeant une structure plus globale du discours où des relations peuvent exister entre des énoncés non adjacents, souvent de nature sémantique ou pragmatique.

3.6 Approches sémantiques

Il est également possible de faire intervenir plus directement les aspects sémantiques dans le processus de résolution des anaphores par l'intermédiaire de connaissances sémantiques ou de représentations sémantiques des énoncés. Par exemple, tout énoncé analysé est traduit sous forme d'une structure sémantique (rangée dans un graphe), et les anaphores sont résolues en utilisant des opérations de simplification, de projection, etc. sur les graphes construits. A l'aide de caractéristiques que l'on peut déduire de la référence anaphorique,

on exploite la structure sémantique du dernier énoncé pour trouver un élément correspondant. D'autres approches comme la DRT [25] permettent de construire un contexte global (contrairement à la théorie de centrage), composé de l'ensemble des référents du discours. Elles proposent une modélisation dynamique de la compréhension d'un discours qui passe par la construction d'une représentation discursive d'un texte permettant de faire un certain nombre de prédictions sur l'interprétation de la suite du discours.

3.6.1 Traitement fondé sur les Graphes Conceptuels

Les graphes conceptuels ont été conçus pour représenter la sémantique du langage naturel ; ils ont évolué pour fournir des systèmes complets au sens de la logique. De façon générale, un graphe conceptuel est défini comme un graphe qui a deux sortes de nœuds :

- les nœuds concepts qui représentent des entités, des attributs, des états, des événements...
- les nœuds relations conceptuelles qui symbolisent les liens qui existent entre deux concepts.

$[CONCEPT] \longrightarrow (RELATION) \longrightarrow [CONCEPT]$ Un arc entrant relie un concept à une relation conceptuelle, et un arc sortant relie une relation conceptuelle à un concept. Ainsi, un graphe conceptuel est orienté, fini[11], connexe[12] et bipartie[13].
Ce modèle présente un grand avantage dans la résolution des ambiguïtés linguistiques rencontrées en langage naturel telles que les problèmes de la polysémie, de la synonymie et de l'anaphore. En outre, il offre un cadre théorique permettant de représenter les constructions sémantiques profondes des verbes et des phrases en langage naturel.

En ce qui concerne le traitement des anaphores, la recherche de l'antécédent, et comme nous avons déjà vu dans les algorithmes de Hobbs et de Lappin et Leass, impose des contraintes de type syntaxique. Par exemple, la prise en compte du genre, du nombre, etc. permettrait de rejeter toute coréférence entre « Paul » et « elle » dans l'exemple 27 :

Exemple 27
Paul et Isabelle vont à l'école. *Elle* a un cartable lourd.

Cependant, il existe des types d'anaphores pour lesquels le recours aux connaissances sémantiques et pragmatiques sur le discours est nécessaire pour leur résolution, les connaissances syntaxiques n'étant pas suffisantes. La théorie de graphes conceptuels fournit une technique élégante pour représenter les références anaphoriques liées aux concepts. Par exemple, dans la phrase 28 :

[11] tout graphe dans une mémoire d'un ordinateur ne peut avoir qu'un nombre fini de noeuds.
[12] deux parties non connectées donnent deux graphes conceptuels.
[13] il ne possède que deux sortes de noeuds : les concepts et les relations conceptuelles ; chaque arc reliant une sorte de noeud à l'autre sorte de noeud.

Exemple 28
Jean téléphone à Marie. Il lui souhaite la bienvenue.

On a deux références anaphoriques *il* représente *Paul* et *lui* représente *Marie*. Pour résoudre ce cas, on indique les références anaphoriques par une étoile suivie par une variable telle que ***X**. Les deux phrases précédentes peuvent être représentées par les graphes de la Figure 1, dans lequel la relation AGT (agent) représente l'entité intervenant de façon active et

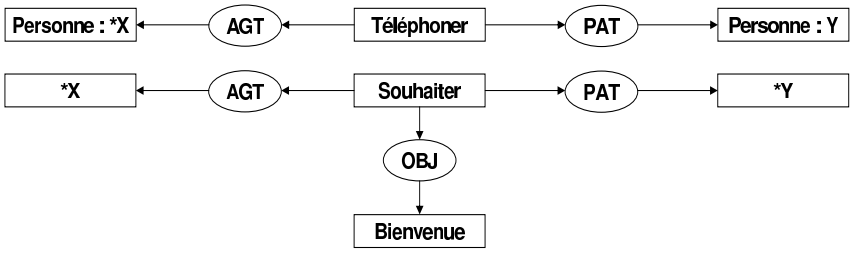


FIG. 1 – Graphes Conceptuels représentant les ER de l'exemple 28

directement dans le procès, la relation PAT (patient) représente l'entité intervenant de façon passive dans le procès et la relation OBJ (objet) représente l'entité affectée par le procès.

3.6.2 Théorie des représentations discursives (DRT)

L'interprétation sémantique d'un énoncé est quelque fois considérée comme un processus qui consiste à intégrer le contenu d'une phrase ou d'un paragraphe à un ensemble de niveau supérieur (paragraphe ou texte) c'est-à-dire au co-contexte existant. Le contexte dans la DRT est représenté comme un ensemble des structures de représentation de discours (DRSs), il sauvegarde les informations véhiculés par le discours. Une DRS est formée de deux composantes : l'univers du discours avec des variables pour les objets et les événements introduits par le discours dŠune part et des conditions portant sur ces variables, d'autre part.

La mise à jour d'une DRS s'effectue de façon dynamique. Des règles d'introduction de nouveaux référents et de conditions sont appliquées en fonction de la structure syntaxique de tout nouvel énoncé. Un indéfini est par exemple considéré comme introduisant une nouvelle variable et une nouvelle condition, correspondant à la prédication introduite par la tête nominale sur cette variable. Pour résoudre une référence pronominale, on introduit une nouvelle variable qui sera liée, par une condition d'égalité, à un référent existant dans le discours, à condition que celui-ci soit accessible et approprié. Les contraintes d'accessibilité sont modélisées en fonction de la structure syntaxique des énoncés. Ainsi, un référent introduit sous la portée d'un opérateur de négation ou dans une construction conditionnelle ne sera plus pleinement accessible pour la suite du discours. Les prédictions de la DRT sur l'interprétation référentielle et plus particulièrement sur l'interprétation pronominale, sont

donc plutôt formulées en termes d'inaccessibilité qu'en termes d'accessibilité. En effet, parmi les référents accessibles sur critères syntaxiques, la DRT n'indique pas comment choisir le référent le plus approprié [35].

Illustrons cette représentation par l'exemple (tiré de [31]) : « John loves Lisa ». Son diagramme DRS est montré dans la figure 2.a.
John(x) et Lisa(x) sont des référents de discours. La DRS de la figure 2 correspond séman-

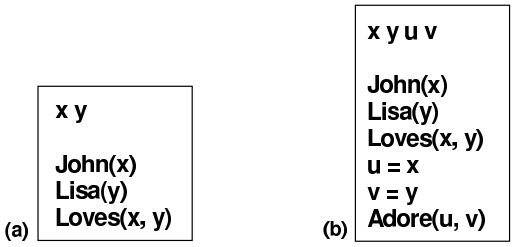


FIG. 2 – Instance d'une DRS

tiquement à la formule logique de premier ordre suivante :

$$\exists x, y John(x) \wedge Lisa(y) \wedge loves(x, y)$$

De même le discours « John loves Lisa. He adores her » est représenté par le diagramme DRS de la figure 2.b.

L'introduction d'un nouvel référent à la DRS se fonde sur la structure syntaxique de l'énoncé, ce qui rend cette approche insuffisante pour un modèle de résolution de la référence dans le contexte de l'oral. En effet, les structures syntaxiques de l'oral peuvent ne pas respecter les normes définies pour l'écrit [2].

3.7 Approche multi stratégies

Les méthodes et algorithme présentés ci-dessus montrent la complexité du phénomène anaphorique, et l'insuffisance des approches purement syntaxiques et les inconvénients de celles syntaxico-sémantiques. Carbonell [7] montre la nécessité d'approches explicitement multi-stratégiques pour la résolution des anaphores. En effet, les informations syntaxiques jouent un rôle principal dans la recherche des antécédents des anaphores intraphrastiques, par contre dans le cas des anaphores interphrastiques, les informations sémantiques et pragmatiques sont indispensables. La résolution des anaphores pourrait être ainsi facilitée par une combinaison de plusieurs stratégies nécessitant l'accès et l'intégration de toutes sortes des connaissances nécessaires pour l'interprétation de discours.

La méthode de résolution proposée consiste d'abord à appliquer les contraintes de sélection pour réduire le nombre des référents candidats, puis à appliquer éventuellement les règles de préférence sémantique sur les candidats restants. Le moteur de résolution a été développé dans un contexte du projet d'analyseur universel[14] [38], qui unifie les connaissances syntaxiques et sémantiques pour produire une analyse complète pour chaque phrase. La résolution des anaphores opère sur l'ensemble des schémas sémantiques instanciés et des arbres syntaxiques. La préférence utilise une méthode de vote pour déterminer le candidat préféré. Enfin, l'antécédent élu sera unifié avec le référent pour restreindre le nombre de cas dans les phrases suivantes ; par exemple si *docteur* est élu comme étant l'antécédent de *elle*, toutes les prochaines anaphores au *docteur* seront obligatoirement de genre féminin ou inconnu.

Nous montrons ci-dessous l'exécution de cette méthode sur deux exemples, en anglais, extraits de [38].

Phrase 1 : The doctor gave John a glass of water

(SENT1
(IS-A *GIVE) (:TIME *PAST) (:AGENT *DOCTOR)
(:OBJECT OBJECT1) (:RECIPIENT *JOHN))
(*DOCTOR
(IS-A *PERSON)) *[unknown gender]*
(OBJECT1
(IS-A *DRINKING-WATER) (:AMOUNT GLASS 1))
(*JOHN
(IS-A *PERSON)(:GENDER M) (:NUMBER *SINGULAR))

frame = (:RECIPIENT *JOHN)
No referents for definite NP
frame = (:OBJECT OBJECT1)
No referents for definite NP
frame = (:AGENT *DOCTOR)
No referents for definite NP

Phrase 2 : John drank it *[it = glass of water]*

(SENT2
(IS-A *INGEST-FOOD) (:TIME *PAST) (:AGENT *JOHN) (:OBJECT OBJECT2))
(*JOHN
(IS-A *PERSON) (:GENDER M) (:NUMBER *SINGULAR))
(OBJECT2

[14]Universal Parser (UP) project.

(IS-A *LIQUID) (:PRO +) (:NUMBER *SINGULAR))

frame = (:OBJECT OBJECT2)
Candidates : ((1 :AGENT *DOCTOR) (1 :AGENT *OBJECT1) (1 :RECIPIENT *JOHN))
after pre-post-cond : ((1 :AGENT *DOCTOR) (1 :OBJECT OBJECT1) (1 :RECIPIENT *JOHN))
after local constr : ((3 :OBJECT OBJECT1))
after case-role constr : ((3 :OBJECT OBJECT1))
referent : (:OBJECT OBJECT1)

frame = (:AGENT *JOHN)
Candidates : ((1 :AGENT *DOCTOR) (1 :AGENT *OBJECT1) (1 :RECIPIENT *JOHN))
after pre-post-cond : ((1 :OBJECT OBJECT1) (1 :RECIPIENT *JOHN))
after NP agreement : ((9 :RECIPIENT *JOHN)) after local constr : ((12 :RECIPIENT *JOHN))
after case-role constr : ((12 :RECIPIENT *JOHN))
referent : (:RECIPIENT *JOHN) *[both JOHN's are coreferential]*

3.8 Conclusion

Cette revue montre les difficultés du traitement de l'anaphore pronominale et les efforts consentis jusqu'à présent pour améliorer le taux de réussite de la recherche des référents. Les approches purement syntaxiques semblent avoir atteint leurs limites notamment en présence d'ambiguïté. On note donc une forte tendance qui plaide pour l'utilisation de toutes les sources de connaissances possibles : syntaxiques, sémantiques, applicatives, encyclopédiques, etc.

4 Prise en compte des descriptions définies

Nous présentons dans le paragraphe suivant les apports des travaux de Vieira et Poesio dans la classification et le traitement des descriptions définies.

4.1 Approche de Vieira et Poesio

Les travaux les plus importants sur les descriptions définies ont été réalisés par Vieira et Poesio [40]. Avant de mettre au point un système de traitement des descriptions définies dans les textes journalistiques, ils se sont intéressés au problème d'annotation des corpus (détection des phénomènes référentiels) par des annotateurs humains, et donc aux problèmes de détection et d'usage. Un des premiers constats a montré des désaccords assez importants entre annotateurs en particulier sur le choix de l'antécédent et fait émerger des questions telles que : sur quels éléments se fonde un interlocuteur pour résoudre une expression référentielle ? quel est le rôle de l'antécédent ? Une première expérience montre que les descriptions définies ne sont pas seulement anaphoriques ; environ dans la moitié des cas

elles sont utilisées pour introduire de nouvelles entités dans le discours (*première mention*) sans un antécédent explicite dans le co-texte.

Une deuxième expérience plus approfondie des différents usages a été réalisée en regroupant de manière différente les phénomènes. Les résultats en terme de fréquence d'apparition sont les suivants :

- 43% à 45% sont utilisés comme expressions *coréférentielles* (comprend toutes les reprises avec ou sans la même tête) ;
- 20% à 25% sont de type *situation plus large* ;
- 18% à 26% sont de type *utilisation non familière* ;
- 6% à 11% sont de type anaphore associative ;
- 0% à 6% de doute.

Une interprétation de ces chiffres pourrait être la suivante ([29]) : « le problème des descriptions définies n'est peut être pas d'identifier leur antécédent, mais d'être sûr qu'elles ont un antécédent ». Dans le cadre de notre système, nous souscrivons entièrement à cette interrogation d'autant plus que dans notre contexte la reconnaissance de la parole génère de l'incertitude.

Vieira et Poesio proposent un système (architecture et méthode) qui permet de détecter et de traiter les descriptions définies. Ils s'appuient sur les fréquences d'apparition des différents phénomènes. Le résultat de l'expérimentation concernant la fréquence des descriptions définies de type *première mention* a mené à développer une méthode heuristiques pour les identifier. Cette méthode est fondée sur des caractéristiques lexicales et syntaxiques des énoncés et/ou des groupes nominaux. Par exemple, la présence de certains adjectifs (comme *best, first, ...*) accompagnés de relatives, de mots prédicats (fact, result, conclusion, ...) complétés par des modifieurs ou encore la présence de déictiques temporels (*day, week, month,...*) permettent de déclencher des filtres et de détecter les DD en première mention. Il faut noter que ces caractéristiques ont été déterminées manuellement. A l'évaluation le taux de rappel est de 69% et le taux de précision est de 72%.

En ce qui concerne les descriptions définies associatives, il y a deux sous-tâches principales dans la résolution : la première est la recherche de l'élément *(ancrage)* dans le texte auquel la description est reliée et la seconde consiste à identifier la relation *(lien)* entre la description et son *ancre*. On distingue les classes suivantes [34] :

- les cas fondés sur des relations lexicales bien définies comme la *synonymie*, l'*hypernoymie* et la *méronymie* qui peuvent être trouvées dans une base de données lexicales comme WordNet (présentée et utilisée dans l'expérimentation de Vieira [40]) ;
- la description définie *nom commun* pour laquelle l'antécédent est un nom propre ; sa résolution nécessite la reconnaissance du type de l'entité dénotéepar ce nom propre ;

- les cas dans lesquels l'antécédent de la description n'est pas le nom principal d'un GN mais un nom modifiant l'antécédent ;
- les cas dans lesquels l'antécédent n'est pas introduit par un GN mais par un GV (*réification*) ;
- les descriptions dont l'antécédent n'est pas mentionné explicitement dans le texte, mais est disponible implicitement car c'est un thème du discours (*industrie* dans un texte sur des compagnies pétrolières) ;
- les cas dans lesquels les relations avec l'antécédent sont fondées sur des connaissances plus générales du type cause-conséquence.

Le dernier type de description définie classée par Vieira et Poesio est l'*anaphore directe*. La stratégie pour résoudre ce cas de description est la suivante : le référent est l'antécédent potentiel pour lequel la tête s'accorde avec le nom principal de la description définie. Deux sous-tâches sont nécessaires :

- identifier l'antécédent potentiel ;
- vérifier l'accord entre antécédent et description définie.

Une solution heuristique a été mise en œuvre [40] pour éviter une analyse syntaxique complète nécessaire pour un tel traitement et pour segmenter le texte afin de limiter la recherche.

Les résultats de Vieira et Poesio sont intéressants à plus d'un titre. Les fréquences d'apparition des différents phénomènes référentiels peuvent être exploités dans un premier temps. Elles ont été calculées sur des corpus d'un genre particulier et demandent sans doute à être recalculées pour d'autres contextes. Les solutions utilisées pour la résolution des références sont complexes et démontrent bien d'une part la difficulté du problème et d'autre part l'intérêt de l'utilisation de nombreuses sources de connaissances.

5 Modélisations et traitements dans un contexte multimodal

La résolution des références multimodales consiste à découvrir le référent d'une expression dans une modalité en utilisant de l'information présente soit dans la même modalité soit dans d'autres modalités. La résolution revient donc à trouver les relations de coréférence entre les termes de différentes modalités en prenant en compte les contextes immédiats ou les historiques des activités sur les différentes modalités. En général, les modélisations mises en œuvre prennent appui sur la langue naturelle et instaurent des liens sémantiques avec les représentations des éléments sur les autres modalités.

Nous allons décrire des travaux qui traitent principalement des liens entre la linguistique et la notion de spatialité. Les approches du problème sont disparates dans la mesure où elles

privilégient chacune un point de vue particulier : le fonctionnement des termes linguistiques (travaux de Vandeloise), les aspects spatiaux avec l'organisation des éléments dans l'espace et l'intervention de la géométrie ainsi que la prise en compte de l'organisation des scènes ou encore le traitement du geste et sa signification mode dans des contextes visuels et linguistiques.

5.1 Modèle de Vandeloise

En situation multimodale, dans laquelle la notion d'espace est importante (soit par la position physique des interlocuteurs -par exemple, demande d'itinéraire dans la rue-, soit par des représentations accessibles visuellement -exemple affichage de carte dans Géoral-), la modalité langagière apparaît particulièrement utile pour exprimer des orientations et des repérages. Il existe donc de nombreux travaux en linguistique sur les relations entre le langage et le cadre spatial. Nous nous intéressons dans ce qui suit aux travaux de Vandeloise [39].

Vandeloise fournit une description fonctionnelle des prépositions spatiales françaises. La cadre théorique de cette approche se fonde sur la notion de ressemblance de famille et utilise les notions de *cible* et de *site* en prenant en compte des aspects fonctionnels du monde qui nous entoure. Après avoir présenté la notion de ressemblance de famille et les aspects fonctionnels, nous en viendrons aux propositions de la modélisation proprement dite.

5.1.1 Notion de ressemblance de famille

Cette notion est définie comme suit : une famille est un concept représenté par la combinaison des traits qui le caractérisent, sachant que chaque trait n'est pas toujours nécessaire ni suffisant. Il s'agit d'une structuration qui permet aux membres d'une catégorie d'être reliés les uns aux autres, sans avoir une propriété en commun définissant la catégorie.

On trouvera deux illustrations de cette notion d'une part à la figure 3(a) qui représente quelques exemplaires de la catégorie « oiseau », et d'autre part, à la figure 3(b)[15] qui vérifie bien la propriété que chaque exemplaire de la catégorie partage au moins une propriété avec un autre membre de la catégorie (exemples extraits de [36]).

5.1.2 Aspects fonctionnels

Une composante fonctionnelle est définie comme *« l'ensemble des connaissances extra-linguistiques de l'espace que partagent les locuteurs d'une même langue »*. Les aspects fonctionnels sont nécessaires pour compléter la géométrie et la logique et doivent donc participer à une bonne modélisation des usages spatiaux de la langue. Cette idée peut être illustrée

[15] A titre d'illustration, *a*, *b*, *c*, *d*, *e*, et *f* peuvent par exemple désigner les membres d'une famille. L'individu *a* présentant par exemple la même couleur de cheveux que l'individu *b*, ce dernier présentant à son tour une ou plusieurs similitudes avec l'individu *c*, etc.

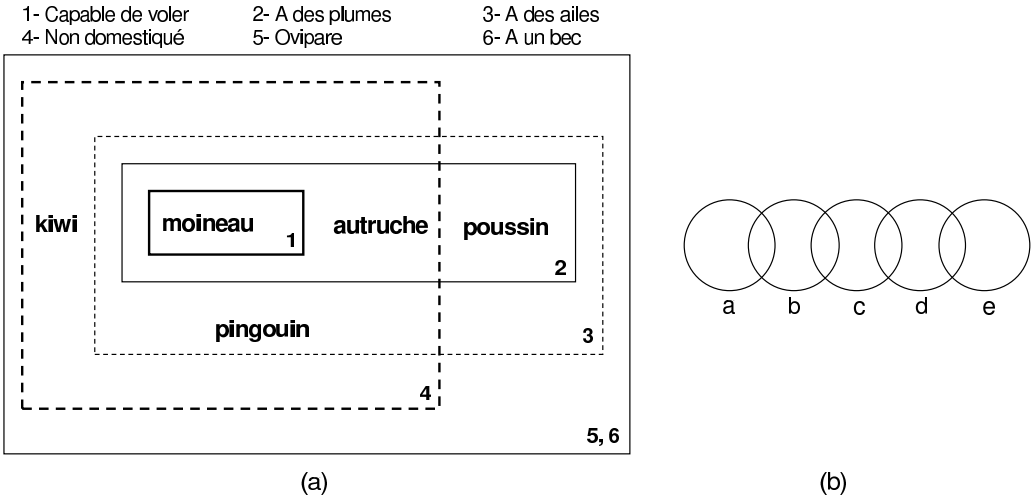


FIG. 3 – La notion de ressemblance de famille

avec les phénomènes de compositionnalité et de transitivité de la logique.

La logique se veut nécessairement compositionnelle, c'est-à-dire que la valeur d'une séquence d'éléments se déduit strictement d'une part de la valeur de ses éléments, d'autre part de ses règles de combinaison qui leur sont applicables. Par contre ce principe ne peut pas être appliqué sur tous les énoncés, certains énoncés expriment beaucoup plus d'information que l'on pourrait en déduire a priori d'une simple combinaison de chacun des termes pris isolément. Ce fait général apparaît également au niveau des énoncé « spatiaux ». Ainsi l'énoncé de l'exemple 29 implique plus que la simple proximité de deux éléments considérés, il indique en plus que Paul se trouve vraisemblablement assis à son bureau et qu'il y travaille.

Exemple 29
Paul est à son bureau

Tout comme la compositionnalité, la transitivité peut provoquer quelques problèmes en modélisation du langage naturel. Elle n'est pas toujours respectée pour la préposition *sur* comme le démontrent les énoncés 30.(a) et 30.(b) qui ne permettent pas de déduire naturellement par transitivité l'énoncé 30.(c).

Exemple 30
Le livre est sur la table. (a)
La table est sur le sol. (b)
Le livre est sur le sol. (c)

Ces différents points montrent qu'avant de procéder à une formalisation de l'espace (qui serait alors trop « brute ») il est nécessaire de préciser le fonctionnement précis des opérateurs spatiaux du langage naturel. Une modélisation fine des aspects spatiaux passe par une nécessaire prise en compte des aspects fonctionnels comme nous avons vu pour les problèmes de compositionnalité et de transitivité. Ce paragraphe et le précédent justifient donc l'importance d'une étude linguistique fine qui tienne compte de tous les niveaux d'analyse, dont la pragmatique. Nous allons étudier le modèle de Vandeloise qui est en mesure d'apporter des éléments de réponse.

Vandeloise fait reposer sa description de la sémantique des prépositions spatiales sur un certain nombre de principes décrits ci-dessous.

Principe extra-linguistique : un objet dont la position est incertaine ne peut pas être localisé sans référence à une entité dont la position est mieux connue.

Vandeloise, en se fondant sur le principe extra-linguistique, définit les notions de *cible* et de *site*. Tandis que la cible joue le rôle d'entité localisée, le site joue celui d'entité localisatrice.

Principe de fixation : un objet peut être qualifié par rapport à sa position usuelle, même si sa position réelle diffère au moment de l'énonciation.
C'est le cas des propositions *« le dessus de »* et *« le dessous de »* appliquées à une bouteille, le dessous de cette dernière ne change pas lorsqu'on la renverse, ainsi les deux propositions sont dites « fixées » par la position normale de la bouteille. On parlera alors d'*orientation intrinsèque* pour ces objets.

Principe de transfert : le locuteur a la faculté de se déplacer en tout point utile pour modifier la perspective de vision de la scène « objective » qu'il décrit[16].

Vandeloise mentionne le fait que si les positions du locuteur sont en nombre illimitées, la conversation, elle, ne fonctionne de manière satisfaisante qu'aussi longtemps que l'interlocuteur est capable de suivre la trajectoire mentale du locuteur. Cette trajectoire mentale partagée s'inscrit dans le contexte plus général de la pragmatique dans le sens où le discours a un rôle essentiel à jouer dans l'interprétation d'un énoncé.

Illustrons le modèle de Vandeloise à partir des prépositions spatiales *« sur / sous »*, pour lesquelles il énonce la règle d'usage **S** suivante :
S : A est « sur / sous » B si sa cible est le deuxième / le premier élément de la relation porteur / porté et son site, le premier / le deuxième élément de cette relation.

La relation porteur / porté se comporte comme une ressemblance de famille dont les diffé-

[16] Les notions de perspective et de scène objective seront reprises avec la notion de cadre pour laquelle une rupture de cadres entre deux locuteurs sera généralement source d'ambiguïté.

rentes caractéristiques sont les suivantes :

- Caractéristique A : si A est *sur / sous* B, la cible est généralement[17] *plus haute / plus basse* que le site.
- Caractéristique B : si A est *sur* B, il y a généralement un contact direct entre la cible et le site.
- Caractéristique C : si A est *sous* B, la cible est généralement rendue inaccessible à la perception par le site.
- Caractéristique D : dans les relations A est *sur / sous* B, la cible est généralement plus petite que le site.
- Caractéristique E : si A est *sur* B, l'action du site s'oppose à l'action de la pesanteur sur la cible.

L'un des points forts de ce travail réside dans la clarté avec laquelle les concepts fondamentaux sont examinés et mis en évidence. Cependant l'une des limites de l'approche réside dans le fait qu'elle n'est pas directement implémentable. Par exemple le concept de ressemblance de famille qui paraît naturel est difficilement formalisable entièrement et donc difficilement programmable.

5.2 Approche géométrique

Bien que la géométrie semble bien adaptée à la modélisation de l'espace, elle ne peut pas venir à bout des nombreux problèmes qui concernent la modélisation des usages spatiaux de la langue [36]. D. Schang a tenté d'illustrer ce fait au moyen de certains usages des prépositions ; nous présentons le cas des prépositions *près de* et *loin de* dans les phrases de type *« A est près de B »*. La méthode qui semble la mieux adaptée pour situer les objets les uns par rapport aux autres est la distance euclidienne classique.

∀ A et B deux objets, la distance séparant A et B est la plus courte des distances entre un point $A_i \in A$ et un point $B_j \in B$ donc $d(A,B) = Min_{i,j}\, d(A_i,B_j)$.

L'usage des prépositions *près de* et *loin de* est difficile à formaliser à l'aide de cette seule distance euclidienne. Prenant l'exemple d'un coureur cycliste qui pourrait dire *« Paris est très loin »* alors qu'un aviateur serait peut-être amené à dire *« Paris est très près »*. Ces exemples montrent qu'il faut prendre en compte dans toute modélisation d'une situation de dialogue le *locuteur* ainsi que son *interlocuteur*. En reprenant l'exemple de l'aviateur, dans le contexte où sa réserve de carburant est quasi nulle, ce qui risque de le faire tomber en panne d'une seconde à l'autre, il paraîtrait alors très naturel qu'il dise *« Paris est encore loin »*. Donc l'outil géométrique ne semble pas, seul, pouvoir fournir de description assez fine des mécanismes mis en œuvre dans les énoncés traitant de l'espace. Des chercheurs ont proposé la prise en compte d'éléments contextuels dans le cadre d'approches utilisant la géométrie.

[17] Le terme *généralement* est précisé pour des exemples du type : *la mouche est sur le plafond.*

TAB. 4 – Les relations topologiques

Relation entre A et B	$\bar{A} \cap \bar{B}$	$\dot{A} \cap \dot{B}$	$\bar{A} \cap \dot{B}$	$\dot{A} \cap \bar{B}$
disjoint	$\neg\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
tangent	$\neg\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
chevauche	$\neg\emptyset$	$\neg\emptyset$	$\neg\emptyset$	$\neg\emptyset$
contient sur un bord	$\neg\emptyset$	$\neg\emptyset$	$\neg\emptyset$	$\emptyset$
inclus sur un bord	$\emptyset$	$\neg\emptyset$	$\neg\emptyset$	$\neg\emptyset$
contient	$\neg\emptyset$	$\neg\emptyset$	$\emptyset$	$\neg\emptyset$
inclus	$\emptyset$	$\neg\emptyset$	$\emptyset$	$\neg\emptyset$
égal	$\neg\emptyset$	$\neg\emptyset$	$\emptyset$	$\emptyset$

5.3 Modèle d'Hernández

C'est une approche qualitative plutôt que quantitative. Elle s'adapte au contexte comme nous allons le voir. Hernández [22] propose un modèle adapté à la prise en compte de la projection 2D d'une scène 3D. L'auteur précise que seulement deux facteurs suffisent à déterminer la position relative d'un objet par rapport à un autre : l'orientation des objets et leur extension. Il en résulte la définition de deux classes de relations spatiales disjonctives qui permettront de localiser précisément la position d'un objet par rapport à un autre à l'aide d'un doublet relation : topologique et d'orientation.

5.3.1 Relations topologiques

Huit relations topologiques binaires ont été détaillées : disjoint, tangent, chevauche, contient sur un bord, inclus sur un bord, contient, inclus et égal. Ces différentes relations sont spécifiées dans le tableau 4 où $\bar{A}$ désigne l'extérieur de A au sens topologique du terme et $\dot{A}$ désigne l'intérieur de A.

5.3.2 Relations d'orientation

Hernández considère tout d'abord des objets sans extension et propose de projeter sur l'objet de référence (appelé site chez Vandeloise) l'un des systèmes d'axes de la figure 4.

Imaginons que l'on souhaite localiser deux cibles C_1 et C_2 par rapport à un site donné ; si on obtient la même préposition pour localiser les deux cibles au niveau i, il suffit de passer au niveau $i+1$ et ainsi de suite. Si, arrivé au niveau 3, il y a encore ambiguïté, Hernández pense que plutôt qu'affiner encore le système d'axes, il convient de changer le site. Cette approche permet donc de prendre finement en compte le contexte en ne travaillant qu'au niveau de granularité nécessaire et suffisant [36].

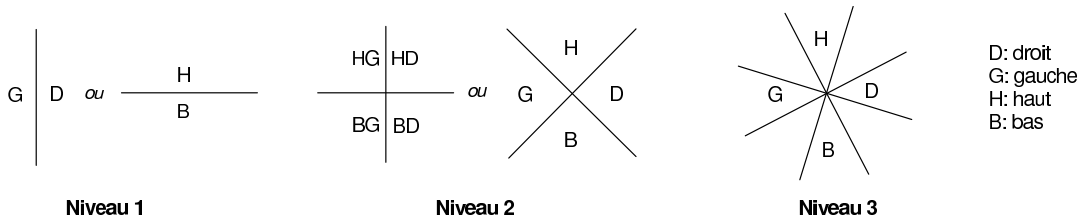


FIG. 4 – Le modèle de Hernández

Dans le cas des objets avec extension, Hernández démontre tout d'abord qu'on ne peut pas se contenter des centres de gravité pour décrire précisément la position d'un objet A par rapport à un objet B. Il suffit de prendre l'exemple de la figure 5 (tirée de [36]) pour lequel si nous prenons seulement en considération les centres de gravité nous aboutirons à situer B comme étant à droite de A dans les deux cas de figure, ce qui est faux dans le premier cas. Pour résoudre ce problème Hernández distingue trois cas selon que la cible et le site :

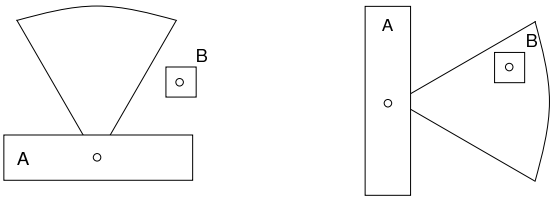


FIG. 5 – Les secteurs d'acceptabilité

- se trouvent *distants* l'un de l'autre
- se trouvent en situation de *proximité*
- se *chevauchent*

Le point faible de cette solution se situe dans le caractère *ad hoc* des mesures de *distance* et de *proximité* qui sont liées directement à l'application. L'auteur propose que la cible est dite à proximité du site si son centre de gravité se situe au dessous de trois fois le diamètre du site. Les algorithmes associés aux trois types de distances considérées pour repérer une cible par rapport un site sont résumés comme suit :

- pour les objets distants, il suffit de tester dans quel secteur se trouve le centre de gravité de la cible ;
- pour les objets en proximité, les zones d'acceptabilité sont étendues pour les côtés et réduites pour les coins. Dans les cas ambigus, Hernández propose de recourir au

système d'axes complémentaire du même niveau *i*, si l'ambiguïté demeure, il propose de revenir au niveau *i-1* et de se poser le même problème ;
- pour les objets qui se chevauchent, le facteur déterminant est la longueur du côté du site qui fait face à la cible. Sans donner de précision l'auteur mentionne cependant que pour les objets étendus tels que les routes et les rivières on peut se contenter des prépositions de niveau 1 comme « à droite de/à gauche de » et « devant/derrière ».

Cette contribution propose de prendre en compte le contexte en ayant recours à trois niveaux de granularité. Le point délicat de l'approche provient du caractère *ad hoc* de certaines valeurs utilisées (l'identification des objets en situation distante ou à proximité) qui met en cause sa généricité.

5.4 Approche fondée sur la notion de cadre (Schang)

En ce qui concerne le raisonnement spatial, les « cadres » de Schang [36] permettent de représenter des structures de boîtes pour gérer des contextes imbriqués hiérarchiquement dans un espace bidimensionnel pour la représentation d'objets d'une interface graphique. Ce modèle est fondé sur celui de Hernández (cf. 5.3) pour le calcul des référents spatiaux comme nous allons voir par la suite. Un des problèmes posés par les prépositions spatiales est qu'elles ne peuvent pas être traitées par la logique du premier ordre [33]. En effet, être « *à droite* » d'un objet, par exemple, n'est pas une propriété purement binaire, vraie ou fausse. Dans l'exemple de la figure 6 (tirée de [33]), pour résoudre l'expression référentielle « *le rond à droite du carré* », il faut prendre en compte la distance entre les objets candidats et le site (le carré X), la distance par rapport à l'axe horizontal. Or si on considère le prédicat « *être à droite de* » uniquement sous son aspect logique, le rond en haut dans la figure 6 pourrait être un aussi bon candidat que le rond entouré d'un trait en pointillé, ce qui n'est pas le cas dans la réalité.

Notons que le modèle de Schang est restreint au cas des références purement langagières dans le pilotage d'un système affichant des objets graphiques (cercle, triangle, carré et rectangle) en deux dimensions en gérant en particulier les références spatiales qui peuvent porter sur ces objets. En plus de la notion de cadre le modèle utilise les notions de saillance (reprise ultérieurement par Landragin [26] et qui fait partie de notre réflexion) et de prototype que nous allons présenter par la suite. La notion de saillance visuelle intervient dans toute communication homme-machine dans laquelle la vision intervient. Ainsi, si l'utilisateur souhaite désigner un ou plusieurs objets, la référence se fera par discrimination sur une ou plusieurs propriétés apparente(s) que possède(nt) l'(les) entité(s) référencée(s) et que ne possèdent pas les autres. La notion de prototype a également été introduite pour remettre en cause la vision classique de la catégorisation : un objet appartient ou n'appartient pas à une catégorie selon le degré de ressemblance qu'il entretient avec un exemplaire particulier de cette catégorie ou prototype.

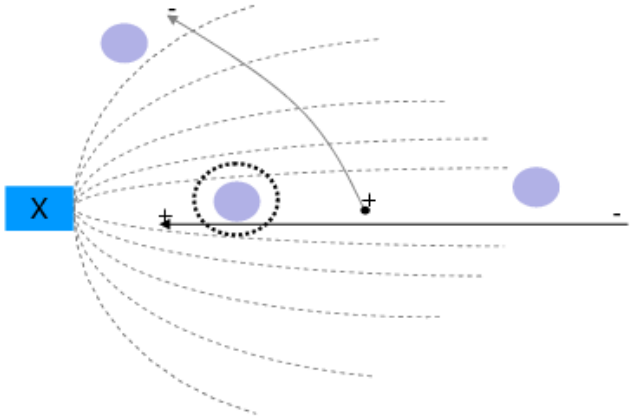


FIG. 6 – Caractéristiques quantitatives en deux dimensions du prédicat référentiel spatial « *à droite de* » dans « *le rond à droite de* ***X*** ».

Le modèle de calcul des référents spatiaux de Schang consiste à étendre progressivement une aire de recherche jusqu'à isoler un objet ou un groupe homogène perceptif (pour le cas où plusieurs objets sont concernés). Ainsi Schang a noté que l'extension de l'aire spatiale de recherche dépend de la préposition spatiale qui est mise en jeu. Pour clarifier ces algorithmes de recherche nous prenons un cas simple (tout d'abord sans la prise en compte de la notion de cadre), la localisation absolue d'une entité unique à l'aide d'une préposition simple comme dans « *détruis le cercle à droite* ». L' algorithme simple proprement dit est le suivant :

tant que (*aucun objet du type recherché ne rencontre l'aire de recherche*) (figure 7.a)
agrandir cette aire à partir de la droite (figure 7.b)

Schang a estimé que la notion de cadre est intéressante pour résoudre des cas ambigus et ainsi pour faire la recherche de la (des) cible(s) sur une partie de l'écran. Selon l'auteur, le cadre est une entité qui restreint l'univers à une portion d'espace ou à la connexion de portions d'espace nécessaires et suffisantes permettant la terminaison du raisonnement spatial courant. C'est en fait une instanciation du concept plus général de contexte. Ainsi le cadre joue le rôle du contexte spatial pertinent.

L'auteur introduit aussi la notion de *focus* qui est la zone à l'intérieur du cadre de référence où se trouve l'entité (respectivement les entités) dont on parle, par opposition aux zones où l'on est sûr qu'elle(s) ne se trouve(nt) pas.

Grâce aux principes de cadre de référence et de focus, des cas plus compliqués pourront être traités. L'exemple « *Agrandis les carrés de droite* » suivi de « *Euh non, rétrécis le* » et

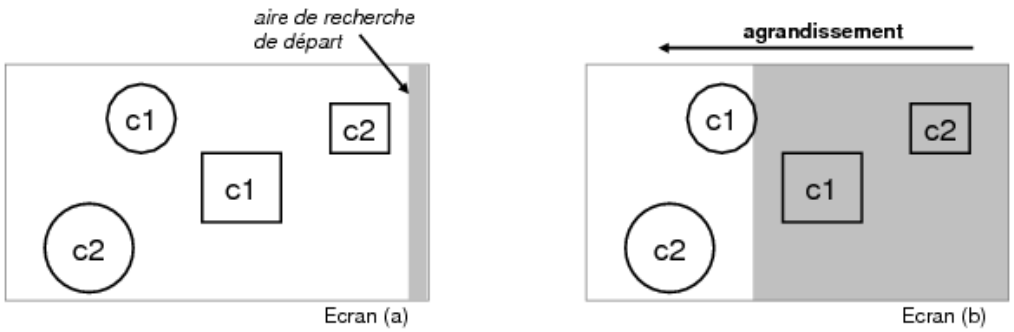


FIG. 7 – Modèle de calcul des référents « absolu » pour la préposition « de droite ».

par la suite « *Détruis l'autre* » (figure 8) pourrait être résolu. Dans ce cas, le suivi discursif du dialogue à l'aide de la notion de cadre de référence (figure 8.b et 8.c) permet tout naturellement de prendre en compte la dernière référence « *détruis l'autre* » sur la base du cadre de référence qui englobe les deux carrés (figure 8.d). Cette approche, comme celle d'Hernández, prend en compte le contexte mais sans avoir différents niveaux de granularité. Ces niveaux de granularité et leurs éventuels sous-niveaux distingués par Hernández ont été fusionnés dans cette nouvelle approche. Ainsi Schang propose une méthode de recherche élégante qui s'adapte en fonction du discours et qui est censée correspondre à la représentation mentale de l'usager.

5.5 Apports de perception visuelle et de la pertinence (Landragin)

La multimodalité oro-gestuelle fait intervenir non seulement les deux modalités d'expression que sont la parole et le geste, mais également le mode de support qu'est la perception visuelle. En effet, le geste démonstratif s'appuie sur le contexte visuel ; sa précision et sa forme dépendent des caractéristiques visuelles de l'objet visé et de sa disposition par rapport aux objets non visés [26].

5.5.1 Objets visuellement saillants

Un objet saillant est un objet qui se distingue des autres objets et qui se trouve mis en valeur [26]. Ce principe peut s'appliquer également à une zone qui se distingue particulièrement de l'ensemble constitué par la scène visuelle. Nous souhaitons résoudre des ambiguïtés de désignations gestuelles grâce à ce type de saillance. En effet, pour un geste qui s'appuie sur le contexte visuel, son empan et sa forme dépendent des caractéristiques visuelles de l'objet visé (par exemple, sa taille) et de sa disposition par rapport aux objets non visés.

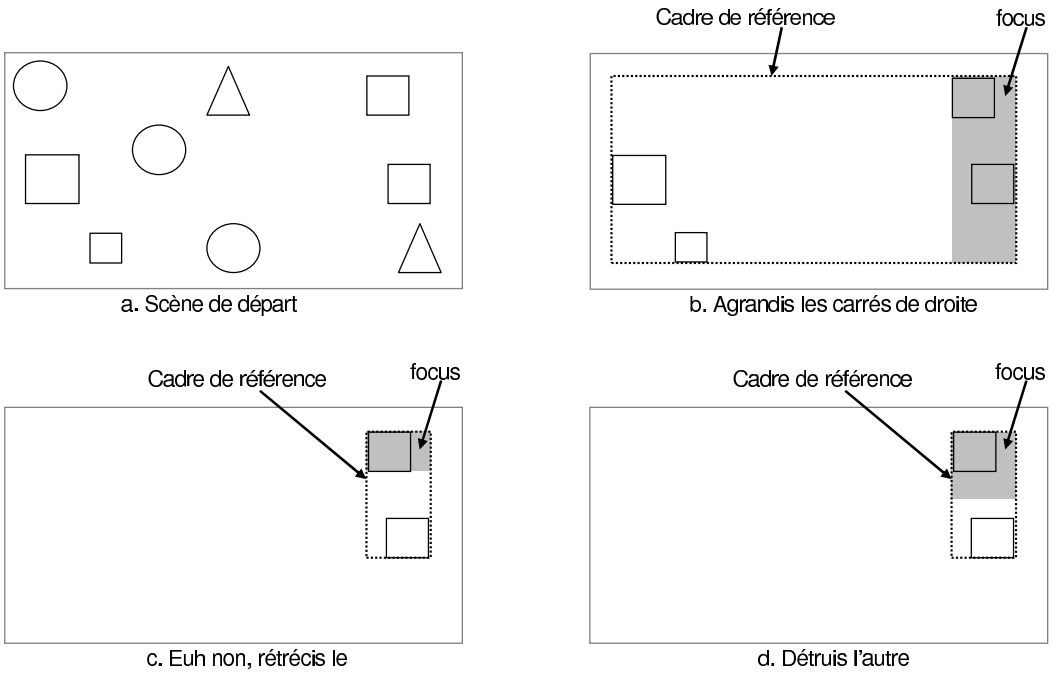


FIG. 8 – De l'intérêt de cadre de référence.

Landragin fait la distinction entre la saillance absolue et la saillance relative : un objet est absolument saillant dans le cas où on peut le distinguer et le mettre en valeur par rapport aux autres entités de la scène visuelle indépendamment de la situation de l'interaction. Il est relativement saillant si cette propriété change de vérité en fonction de l'avancement de l'interaction. L'auteur tient compte aussi de la distinction entre saillance explicite et saillance implicite : le geste ostensif[18] rend le référent saillant et constitue le critère de saillance explicite. En faisant un parallèle avec le langage naturel (saillance linguistique), un énoncé tel que « considère le triangle rouge » rend implicitement le référent (le triangle rouge) saillant pour la suite du dialogue (par exemple, « peins le en bleu »).

5.5.2 Détection des groupes perceptifs et des objets saillants

Il s'agit de structurer en groupes les objets de la scène visuelle. Ces groupes disposent d'une structuration qui permet au système de tenir compte des focalisations spatiales de l'utilisateur et de comprendre les expressions reposant implicitement sur ces focalisations. Les trois principaux critères de groupage sont la proximité, la ressemblance, et la continuité

[18] geste de désignation qui peut prendre appui sur le contexte visuel.

[26]. Les critères de proximité et de continuité se formalisent à l'aide de calculs géométriques sur les coordonnées des objets, et le critère de ressemblance se formalise grâce aux caractéristiques enregistrées des objets. Comme dans l'approche de Hernández, un point faible de cette solution se situe dans le caractère *ad hoc* de mesure de proximité et de continuité.

Un autre aspect important de la perception visuelle est la notion de saillance visuelle : le fait qu'un objet en particulier se distingue des autres objets et attire l'attention de l'utilisateur. Si le système est capable d'identifier à tout moment l'objet saillant dans la scène, il pourra d'une part modéliser une facette de l'attention de l'utilisateur, lui permettant ainsi de prévoir dans une certaine mesure à quel objet celui-ci va s'intéresser, et d'autre part interpréter correctement les actions de référence fondées sur cette saillance. Landragin montre l'existence d'une telle action référentielle par l'exemple suivant : *« enlève le triangle »*, sans geste ni antécédent linguistique, dans le contexte de la figure 9. L'expression référentielle *« le triangle »*, bien qu'ambiguë du fait de la présence de plusieurs triangles, s'interprète facilement comme référent au triangle gris à gauche. Le système identifie l'objet saillant

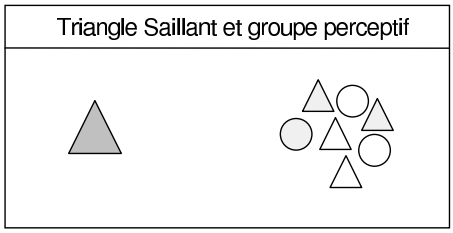


FIG. 9 – Groupement et saillance dans la perception visuelle.

selon plusieurs critères comme la *simplicité* (une forme simple se démarque mieux qu'une forme complexe), la *régularité* et la *symétrie* (une forme ayant une répartition régulière ou symétrique se démarque mieux qu'une forme n'en ayant pas), etc. Ces critères caractérisent une forme qui vient en premier à l'esprit et que nous pouvons considérer comme caractérisant un objet visuellement saillant.

Landragin s'appuie aussi sur la théorie de la pertinence pour le traitement des références aux objets dans le contexte multimodal [26]. L'idée principale de cette théorie est que l'interprétation agit avec deux types de mécanismes, des mécanismes de décodage du message et des mécanismes d'inférence. Ces inférences consistent à déduire du message décodé et du contexte des propositions nouvelles qui constituent l'objet de la communication. Ainsi Landragin a proposé de définir le processus de résolution référentielle dans le cadre dŠun dialogue avec support visuel et gestes de pointage. Il a montré comment la mise en relation de ce qu'expose l'utilisateur et les hypothèses formées par la théorie de la pertinence permettent de sélectionner plus efficacement les bons objets.

L'approche de Landragin met particulièrement en avant l'importance des éléments perceptifs non linguistiques dans le phénomène référentiel pour lequel l'environnement du dialogue ne peut être ignoré, et par conséquent, l'importance de la prise en compte de l'environnement dans l'interprétation des énoncés, principalement vis-à-vis des critères spatiaux de détection des objets saillants et qui sont perçus par l'humain. Comme évoqué précédemment, l'inconvénient de cette approche provient du caractère *ad hoc* de certaines valeurs utilisées (mesure de proximité et continuité).

5.6 Approche de Pineda et Garza

Pineda et Garza [32] présentent une théorie de représentation et d'interprétation de messages multimodaux. Deux modalités sont prises en compte : la modalité langage naturel et la modalité graphique. Le problème traité est celui de la résolution des références multimodales : c'est le problème de recherche de l'antécédent d'un « symbole » dans une modalité en utilisant des informations présentes dans la même ou dans l'autre modalité. L'exemple de la figure 10 (tirée de [32]) montre un message exprimé avec deux modalités différentes, textuelle et graphique, qu'il faut décoder. Pour donner sens aux expressions référentielles *il* et *l'* dans la phrase *il l'a lavé*, il faut les traiter dans leur contexte graphique. Dans ce cas, et après une compréhension des syntagmes *Il*, *l'* et *a lavé*, on peut facilement assigner *il* à l'*homme* et *l'* au *véhicule*. Ce cas peut être assimilé à une inférence anaphorique en considé-

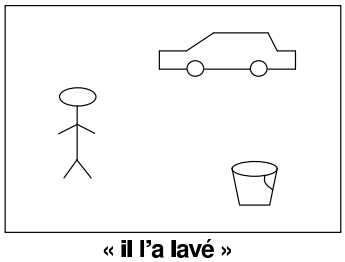


FIG. 10 – Exemple d'une anaphore linguistique avec un antécédent graphique

rant que les informations fournies graphiquement sont exprimées dans le discours suivant : *Il y a un homme, un véhicule et un seau. Il l'a lavé.* Pour traiter le pronom *il* et l'article *l'* à part sans le graphique, la résolution de *il* ne pose aucun problème, par contre pour celle de *l'* il y a deux antécédents possibles, il faut donc des connaissances complémentaires pour pouvoir choisir entre *véhicule* et *seau*, en particulier, le fait que l'homme lave un objet avec de l'eau, et que l'eau est transportée dans un seau.

Considérons la situation inverse présentée en figure 11 (tirée de [32]), cette figure est interprétée comme une carte géographique dans le contexte du texte qu'il la précède. La situation décrite est caractérisée par la présence des anaphores picturales dont les antécédents

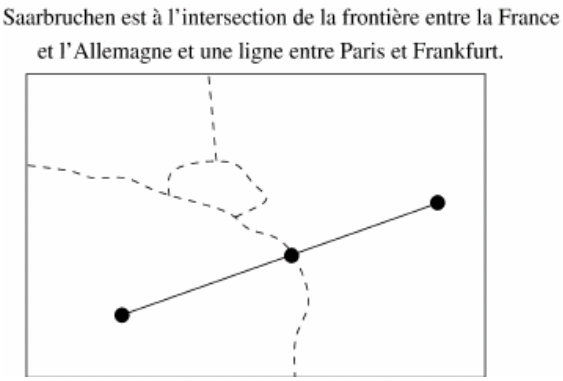


FIG. 11 – Exemple d'une anaphore graphique avec un antécédent linguistique

sont des éléments linguistiques. Ces anaphores sont représentées par des « symboles » graphiques (points et lignes) liés au texte qui a donné du sens à ces symboles. Dans les deux figures, les modalités graphiques et textuelle sont complémentaires, c'est-à-dire qu'il est impossible de traiter le texte de la figure 10 sans le graphique et le graphique de la figure 11 sans le texte. Ces problèmes sont désignés sous le terme d'anaphore graphique. Notons que dans le cas de notre étude, la résolution d'une anaphore graphique est plus compliquée qu'une simple assignation entre les entités de LN et les objets graphiques ou vice versa. La complexité est due au contexte de dialogue entre l'usager et le système et à son lien à la situation spatio-temporelle. En effet un antécédent graphique peut se trouver dans une scène visuelle déjà passée. On peut aussi noter une légère différence entre ce phénomène et la déixis classique pour laquelle le référent se trouve dans l'espace réel courant.

Nous décrivons ci-dessous les principes du modèle proposé par Pineda et Garza et ses caractéristiques.

5.6.1 Description d'un modèle pour la représentation Multimodale

Illustrons ce modèle à partir de la figure 12 (tirée de [32]) qui montre un système de représentation multimodale pour les modalités linguistique et graphique. Ce modèle comprend les composants suivant :

- le cercle **L** contient l'ensemble des expressions en langage naturel ;
- le cercle **G** contient l'ensemble des expressions en langage graphique ;
- le cercle **P** contient l'ensemble des symboles graphiques qui constituent la modalité graphique. Notons que les deux ensembles **G** et **P** sont considérés pour la modalité graphique : les expressions en **G** appartiennent à un langage formel dans lequel la

géométrie des dessins est représentée, alors que **P** contient les symboles graphiques qui ne seront pas manipulés directement ;
– le cercle **W** représente le monde

ρ_{L-G} et ρ_{G-L} représentent les fonctions de transition entre les langages **L** et **G**, et ρ_{P-G} et ρ_{G-P} représentent les fonctions de transition entre les langages **G** et **P**. Lors de la phase d'interprétation graphique, la fonction de transition ρ_{P-G} fournit la représentation des objets de la modalité graphique en des expressions de G. Ainsi la fonction ρ_{G-P} assure la transition des expressions géométriques de G vers les dessins (ex : point, ligne, etc.). Le cercle **W** et les fonctions F_L et F_P constituent un système d'interprétation multimodal. Les paires (**W**,F_L) et (**W**,F_p) constituent respectivement les modèles M_L pour le langage naturel et M_p pour l'interprétation graphique. L'interprétation des expressions de G qui ont

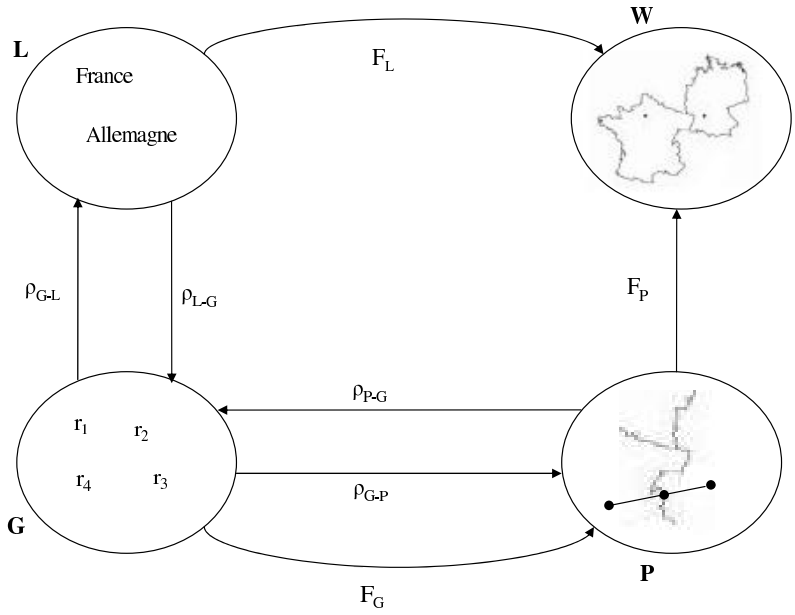


FIG. 12 – Système de représentation multimodal

des relations avec le monde sont définies par la composition $F_L o \rho_{G-L}$ ou, alternativement, par $F_P o \rho_{G-P}$. La paire (**P**,F_G) définit le modèle M_G pour l'interprétation géométrique de G, dans lequel la fonction d'interprétation géométrique F_G assigne une dénotation[19] pour chaque constante de G.

[19]Ces dénotations des constantes de G sont les symboles graphiques.

L'interprétation de l'exemple de la figure 10 (respectivement de la figure 11) au sein de ce système fournit les objets de **L**, **P** et **G**, et la fonction F_L établit les relations entre les constantes linguistiques et les objets. Pour interpréter ces messages multimodaux, il suffit alors d'appliquer la fonction $\rho_{G-P} o \rho_{L-G}$ (respectivement la fonction $\rho_{G-L} o \rho_{P-G}$) pour assigner ainsi *il* et *l'* aux objets graphiques (respectivement les points aux noms de ville) ; il faut calculer, par exemple, $\rho_{G-P}(\rho_{L-G}(il))$, sa valeur est *l'homme* présent sur l'image, et $\rho_{G-P}(\rho_{L-G}(l'))$ donne comme valeur *le véhicule*.

Les expressions référentielles avec ce modèle peuvent être résolues par l'évaluation des compositions de fonctions de transition définies ci-dessus. Cependant les informations du schéma de la figure 12 ne sont pas toutes connues a priori, en particulier les fonctions de transition ρ_{L-G} et ρ_{G-L}. Ainsi l'objectif principal est de calculer ces fonctions, et donc de fournir les relations entre les noms (entités) de L et de G. De plus les informations existantes dans le message multimodal sont habituellement insuffisantes. Il est donc nécessaire de prendre en considération les structures syntaxiques des langages concernés (**L** et **G**), la définition des règles de transition entre ces langages, et les connaissances conceptuelles se rapportant au domaine. Enfin, le modèle doit prendre en compte le problème posé par l'ambiguïté éventuelle des messages multimodaux. Cette ambiguïté peut trouver son origine de manière classique dans le langage naturel, mais aussi dans l'expression graphique. Cette notion d'ambiguïté dans les systèmes multimodaux diffère, selon Pineda et Garza, de celle en langage naturel dans lequel une expression ambiguë possède plusieurs interprétations. Par exemple, la carte peut être représentée par une expression de G qui dénote par transition dans P un simple élément (ex. Europe), ou par plusieurs expressions dans G qui réfèrent une partition dans P (ex. les pays européens). Une meilleure formalisation de la représentation graphique, à l'aide d'une définition d'un langage doté d'une syntaxe et d'une sémantique bien formées, va alors permettre de traiter l'ambiguïté via les relations de transition entre les langages naturel et graphique.

5.6.2 Différents langages

Pineda et Garza [32] présentent la définition détaillée des syntaxes et des sémantiques des langages L, P et G pour résoudre l'ambiguïté multimodale. Nous allons exposer les définitions de ces langages sans entrer dans tous les détails qui ne font pas partie, pour le moment, de notre centre d'intérêt.

Définition du langage L Le langage L contient la partie textuelle des messages multimodaux dans le domaine cartographique, il contient des expressions comme *« Saarbrucken est à l'intersection de la frontière entre la France et l'Allemagne et d'une ligne entre Paris et Frankfurt »*, qui représente la partie langage naturel de l'exemple de la figure 11. Les constantes comme *France* et *Allemagne* et toutes les sous-expressions de la phrase précédente comme *la frontière entre la France et l'Allemagne* sont incluses dans L. De plus L contient des expressions qui sont des connaissances générales nécessaires pour l'interprétation comme

France est un pays, *Frankfurt est une ville en Allemagne* etc. De manière formelle, L comprend les éléments suivants (pour l'anglais) :

- les catégories syntaxiques de base de L sont : *t, IV, ADJ, CN* et *CN'* avec *t* est la catégorie des phrases, *IV* est celle des verbes intransitifs, *ADJ* est la catégorie des adjectifs et *CN* et *CN'* sont respectivement la catégorie des noms communs qui se transforment en prédicats graphiques (ex. : ville, ligne etc.) et celle des noms communs qui se transforment en concepts abstraits (ex. : ouest).
- Si A et B sont deux catégories syntaxiques alors A/B est une catégorie[20].

Les autres catégories traditionnelles du langage naturel comme les verbes transitifs (TV), les termes (T), les phrases propositionnelles (PP) peuvent être dérivés des catégories de base.

Définition du langage P L'objectif de la définition de la syntaxe et de la sémantique du langage **P** est de caractériser la famille des dessins qui peuvent être interprétés comme des cartes, et de discriminer ces dessins d'autres types de configurations graphiques constituées de points, courbes et régions. La formalisation de **P** a pour but d'être capable de parler des cartes comme d'une modalité où une modalité, au sens de Pineda et Garza, est un système de codage pour les symboles exprimés à travers un média et pour lequel un système multimodal de représentation relie d'une façon systématique l'information exprimée à travers différents systèmes de codage. Les types des éléments de P sont point, ligne, courbe, région, zone, région_composée, ensemble_point, ensemble_ligne et carte.

Définition du langage G G contient des symboles qui réfèrent aux objets graphiques et aux configurations d'une part, et qui expriment d'autre part la transition des expressions quantifiées de L. G est un langage formel avec des constantes et des variables, et toute expression bien formée n'a pas de transition dans L. Les transitions utiles portent , par exemple,sur les noms et les descriptions des objets géométriques et des configurations.

Les fonctions de transition entre les différents langages sont formellement définies pour assigner aux éléments d'un langage des éléments dans un autre langage en fonction du message multimodal (cf. l'exemple des figures 10 et 11). Pineda et Garza proposent aussi un algorithme limité à l'interprétation des noms propres et des pronoms dans les messages multimodaux. Notons que l'espace de représentation des objets du monde intégré dans le modèle semble ad hoc. En outre, certains des éléments des différents langages semblent être pré-calculés ce qui peut mettre en cause la généricité de ce modèle.

[20]Une expression de catégorie A/B se joint avec une expression de catégorie B pour donner une expression de catégorie A

5.7 Approche probabiliste

Chai et al. [8] propose une approche probabiliste pour identifier les objets référencés par l'intermédiaire de différents types d'entrée de l'usager dans un système multimodal (entrée orale, geste de désignation sur un écran tactile affichant une carte). L'approche identifie les référents les plus probablement désignés en satisfaisant de manière optimale et conjointe des contraintes sémantiques, temporelles et contextuelles. L'aspect temporel intervient ici car dans leur système il est nécessaire d'aligner temporellement les activités orales et gestuelles de l'usager. Ce problème d'alignement est traité de différentes manières par d'autres auteurs [30] [11] ; il n'intervient pas dans le cas de Géoral, la synchronisation des différentes activités de l'usager étant dirigée par le système par une solution technique simpliste.

la résolution des ER s'effectue avec un algorithme d'appariement de graphes. Les informations issues des différentes modalités d'entrées et des contextes sont codés comme des attributs de graphes relationnels (ARGs), et le modèle de résolution de références revient à apparier de manière probabiliste les différents graphes.

Un ARG consiste en un ensemble des nœuds connectés par des arcs. Chaque nœud représente une entité (une ER à résoudre ou un référent potentiel) ; il est associé à un vecteur de caractéristiques qui encodent les propriétés de l'entité correspondante. Chaque arc représente un ensemble de relations entre deux entités ; il est aussi associé à un vecteur dec caractéristiques qui encodent les propriétés de chaque relation.

Le vecteur de caractéristiques d'un nœud contient les informations sémantiques et temporelles, et celui d'un arc décrit la relation temporelle et sémantique entre des paires d'entités. Une relation temporelle indique l'ordre temporel entre deux entités reliées durant l'interaction qui peut être :

- Précédent : un nœud A précède un nœud B si l'entité représentée par le nœud A est mentionnée avant l'entité représenté par le nœud B dans une modalité spécifique. Par exemple, « this » précède « these two houses » dans l'entrée orale « compare this with these two houses ».
- Concurrent : un nœud A est concurrent avec un nœud B si leurs entités sont référencées ou mentionnées simultanément dans une modalité spécifique. Par exemple, un geste circulaire peut sélectionner un groupe d'objets. Tous les objets sélectionnés sont considérés comme concurrents entre eux.
- Non-concurrent : un nœud A est non-concurrent avec un nœud B si leurs objets/références ne peuvent pas être référencés/mentionnés simultanément et la relation « précédent » ne peut pas être valide entre eux.
- Inconnu : l'ordre temporel entre deux entités est inconnu.

Pour chaque entrée multimodale d'un usager, trois ARGs sont crées :

1. ARG de l'oral (exemple en figure 13) : pour représenter les informations issues de l'entrée orale. Le système utilise le reconnaisseur de parole Via Voice ^{TM} d'IBM ; l'énoncé reconnu est ensuite analysé pour identifier les ER et leurs relations. Chaque ER est identifiée, complétée avec des informations de types morphologique (nombre), syntaxique (pronom, démonstratif,...), sémantique (type du référent visé) et pragmatiques (attributs de l'entité visée : prix, dimension,...) et etiquettée temporellement. Chaque arc entre deux ER représente une relation temporelle et une relation sémantique.
2. ARG du geste : pour représenter les informations issues de l'entrée gestuelle. Les gestes déictiques pris en compte sont le pointé et le dessin de zone (cercle). La reconnaissance du geste va assigner une probabilité à chaque objet censé avoir été désigné. Le calcul de cette probabilité prend en compte le type geste et des considérations de distance entre les objets présents sur l'écran et le geste. Pour chaque geste, un sous-graphe est construit dont un nœud représente un objet sélectionné par le geste. Le vecteur de caractéristiques associé à un nœud va contenir les informations associées à l'objet (type sémantique, attributs pragmatiques, étiquette temporelle, probabilité de sélection). Chaque arc d'un sous-graphe représente une relation sémantique et une relation temporelle. La connexion des sous-graphes permet de représenter l'ARG complet ; cette connexion se fait à l'aide d'arcs de relation temporelle entre gestes.
3. ARG historique : qui représente l'historique de l'interaction. Un ARG historique va contenir la liste des objets qui sont dans le foyer d'attention de l'interaction. On va retrouver dans le graphe les informations qui concernent les objets et leurs relations sémantique et temporelle.

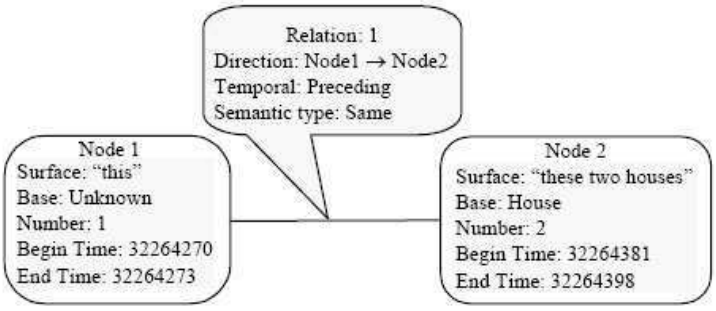


FIG. 13 – ARG de l'oral pour l'entrée « compare this with these two houses »

Ces trois ARGs sont arrangés en deux graphes : graphe référençant et graphe référent. Un graphe référençant G_s est typiquement l'ARG de l'oral. Il contient une collection des ER à résoudre. Un graphe référent G_r est l'agrégation de l'ARG du geste et de l'ARG historique.

Plusieurs hypothèses sont prises en compte pour la résolution des ER :

- une entrée orale peut contenir plusieurs ER qui peuvent faire référence à des objets dans l'historique LN ou à des objets sélectionnés par un geste.
- une ER peut faire référence à un objet qui n'est pas dans l'ARG du geste ou dans l'ARG historique. Comme dans l'exemple « the red house » dans le cas où il n'y a qu'une maison rouge affichée sur l'écran.
- une ER peut faire référence à un groupe d'objets (ex : these houses).

L'algorithme d'appariement de graphe
On note l'ARG référent G_r et l'ARG référençant G_s da la manière suivante :
$G_r = (\{a_x\}, \{r_{xy}\})$, où $\{a_x\}$ est la liste des nœuds et $\{r_{xy}\}$ est la liste des arcs. L'arc r_{xy} relie les nœuds a_x et a_y.

$G_s = (\{a_m\}, \{\gamma_{mn}\})$, où $\{a_m\}$ est la liste des nœuds et $\{\gamma_{mn}\}$ est la liste des arcs. L'arc γ_{mn} relie les nœuds a_m et a_n.

L'algorithme cherche le meilleur appariement possible entre le graphe référent et le graphe référençant par maximisation de la quantité $Q(G_r, G_s) =$

$$\sum_x \sum_m P(a_x, a_m)\xi(a_x, a_m) + \sum_x \sum_y \sum_m \sum_n P(a_x, a_m)P(a_y, a_n)\psi(r_{xy}, \gamma_{mn})$$

$P(a_x, a_m)$ est la probabilité de correspondance entre le nœud référent a_x (du graphe référent) et le nœud référençant a_m (du graphe référençant). $\sum P(a_x, a_m) = 1$ si le nœud de l'ER a_m fait référence à un seul objet. Le terme $Q(G_r, G_s)$ considère la similarité $\xi(a_x, a_m)$ entre les nœuds référents dans les deux graphes et la similarité $\psi(r_{xy}, \gamma_{mn})$ entre les arcs. Ces fonctions de similarité sont déterminées empiriquement à partir d'une série des tests. Notons que la fonction $\xi(a_x, a_m)$ est fondée sur les propriétés de deux nœuds a_x et a_m.

Quand l'algorithme de maximisation de $Q(G_r, G_s)$ converge, la plus grande probabilité $P(a_x, a_m)$ est choisie. Si elle est supérieure à un seuil donné, le système considère que le référent a_x est trouvé pour l'ER a_m. Il y a ambiguïté dans le cas où au moins deux nœuds en correspondance pour a_m sont trouvés et que a_m doit faire référence à un seul objet ; le système pose alors une question de clarification à l'usager.

Cette approche utilise des contraintes sémantiques (fonctions de similarités ξ et ψ), temporelles (l'ordre temporel entre les nœuds : précèdent, concurrent, etc.), et contextuelles (ARG de l'historique) pour traiter les ER multimodales. Ainsi, l'intérêt de cette approche réside dans l'algorithme d'appariement qui intègre ces trois contraintes. Le calcul de similarité entre les nœuds ou entre les arcs semble peu généralisable compte tenu qu'il est fondé sur un corpus et une application particuliers.

5.8 Synthèse sur les traitements en contexte multimodal

On notera tout d'abord que les méthodes présentées (autres que celle de Pineda et Garza et de Chai et al.) ne traitent que la désignation (dans le cadre unimodal et multimodal) d'objets de scène discontinue qui contient des objets géométriques simples :triangle, carré, des cercle, etc. Cela met en cause leur adaptabilité à un système comme Géoral qui contient des objets plus complexes ponctuels, linéaires ou non réguliers. Cette difficulté vient d'une part des relations qui existent entre les objets linéaires (par exemple des similitudes graphiques telles que les routes et les rivières qui ont la même forme, ou des relations du monde réel comme deux rivières qui se rejoignent) et d'autre part, de la manière complètement différente dont ces objets sont perçus et donc désignés par l'usager.

Après avoir présenté les limites des approches d'inspiration purement linguistique ou purement géométrique, nous avons exposé l'approche de Hernández qui est en faveur des approches linguistiques comme cela apparaît dans ses travaux. Ensuite nous avons souligné l'importance de la prise en compte de la notion de cadre avec Schang et celle de la perception visuelle avec Landragin. Nous avons montré la nécessité d'un modèle de résolution des références multimodales avec différents langages de codage des informations et les transitions entre eux comme celui de Pineda et Garza. Enfin nous avons montré l'intérêt de la prise en compte de la sémantique et du contexte dans une approche probabiliste. Nous nous inspirerons de ces deux derniers modèles pour élaborer notre propre modèle.

6 Discussion

Dans le contexte de communications homme machine multimodale, la résolution des expressions référentielles multimodales va nécessiter des connaissances et des traitements propres à chaque modalité ainsi que des connaissances et des mécanismes qui doivent mettre en œuvre simultanément plusieurs modalités.
Concernant la modalité langagière, après la phase de reconnaissance de la parole, l'énoncé doit être analysé pour détecter et typer les ER. Ce traitement doit utiliser les sources de connaissances classiques (morphologiques, syntaxiques, sémantiques) mais aussi pragmatiques liées à l'application et aux contextes historique et courant de l'interaction. Selon le type des ER (anaphore, déictique, première mention), une première proposition de résolution peut être émise lors du traitement. Il est possible par exemple d'émettre l'hypothèse qu'une ER est bien en première mention en fonction des connaissances que le système possède sur le contexte visuel courant ; il est également possible de résoudre linguistiquement une anaphore démonstrative en se réservant le fait de confirmer ou d'infirmer la résolution à l'aide de la modalité gestuelle.

Le traitement du geste consiste dans une première étape en une analyse de la trajectoire gestuelle qui permet de la décomposer à partir de primitives représentant des formes simples (point, arc, polyligne, zone, etc.). Il est ensuite possible d'obtenir les objets censés

être désignés en fonction de l'affichage. Une fois ces traitements unimodaux effectués, il est nécessaire de passer par une phase d'interprétation ou de fusion multimodale durant laquelle les ER multimodales (cf. exemples 5, 6, 7, 8, 9) doivent être résolues ou confirmées. Cette interprétation ne peut être que contextuelle dans la mesure où l'ensemble des connaissances linguistiques, sémantiques, applicatives, situationnelles (instant donné de l'interaction) ainsi que les données performatives des mode oral, haptique et visuel sont partie prenante de ce processus de haut niveau. Les incertitudes inhérentes au langage naturel et celles dues aux performances de l'usager (phénomènes velléitaires à l'oral, hésitation dans la gestuelle) ainsi qu'à celles du système (reconnaissance de la parole, traitement de la langue naturelle) compliquent d'autant plus le processus. Elles vont nécessiter une utilisation des sources de connaissances très intégrée. Enfin rappelons que le modèle à construire vise aussi un certain niveau de généricité pour offrir l'accès à la même application à partir de terminaux mobiles présentant des caractéristiques haptiques et visuelles différentes (PDA et téléphone portable par exemple) et également pour permettre des changements aisés d'applications.

Les idées mise en œuvre dans le système de Pineda et Garza semblent appropriées pour élaborer notre propre modèle étant donné la possibilité de prise en compte des objectifs mentionnés ci-dessus et leurs champ d'application à des données géographiques. Nous travaillons d'une part à adapter les résultats qui nous paraissent les plus intéressants (les différents langages, les possibilités d'inférence) et d'autre part à y intégrer le traitement de la saillance visuelle (travaux de Landragin), celui des contraintes sémantiques (Vandeloise) et contextuelles ainsi que des probabilités notamment pour la désambiguïsation des gestes de désignation.

Références

[1] James Allen. *Natural Language Understanding*. The Benjamin/Cummings Publishing Company Inc., 1987.

[2] Claire Blanche-Benveniste. *Le français parlé : études grammaticales*. Editions du CNRS, Paris, 1990.

[3] Daniel G. Bobrow. A question-answering system for high school algebra word problems. *AFIPS Conference Proceedings*, 26 :591–614, 1964.

[4] Susan E. Brennan. Centering as a psychological resource for achieving joint reference in spontaneous discourse. *In Walker et al*, pages 227–249, 1997.

[5] Susan E. Brennan, Marilyn W. Friedman, and Carl J. Pollard. A centering approach to pronouns. *Computational Linguistics*, 25 :155–162, 1987.

[6] Harry Bunt. Interaction management functions and context representation requirements. In *Actes TWLT 11. S. Luperfoy, A. Nijholt and G. Veldhuijzen van Santen, editeurs, Tilburg University, The Netherlands*, 1995.

[7] James G. Carbonell and Ralf D. Brown. Anaphora resolution : a multi-strategy approach. *COLING*, 1 :96–101, 1988.

[8] Joyce Y. Chai, Pengyu Hong, and Michelle X. Zhou. A probabilistic approach to reference resolution in multimodal user interfaces. In *IUI '04 : Proceedings of the 9th International Conference on Intelligent user interface*, pages 70–77, New York, NY, USA, 2004. ACM Press.

[9] Herbert H. Clark and Deanna Wilkes-Gibbs. *Intentions in Communication*, chapter referring as a Collaborative Process, pages 463–493. The MIT Press, 1990.

[10] Philip R. Cohen. The pragmatics of referring and the modality of communication. *Computational Linguistics*, Volume 10, number 2 :97–146, 1984.

[11] Jacob Eisenstein and Chris Mario Christoudias. A salience-based approach to gesture-speech alignment. In *HLT-NAACL*, pages 25–32, 2004.

[12] Dominique Estival and Francoise Gayral. A study of the context(s) in a specific type of texts : car accident reports. Technical report, LIPN, 1994.

[13] Annie Gal, Guy Lapalme, and Patrick St-Dizier. *Prolog pour l'analyse automatique du langage naturel*. Editions Eyrolles, Paris, 1988.

[14] Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. Providing a unified account of definite noun phrases in discourse , proceedings of the 21st conference on association for computational linguistics, cambridge, massachusetts. *Computational Linguistics*, 21 :44–50, June 15-17 1983.

[15] Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. Centering : a framework for modelling the local coherence of discourse. *Computational Linguistics*, 21 (2) :203–225, 1995.

[16] Barbara J. Grosz and Candace L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12 :175 – 204, 1986.

[17] Marc Guyomard. Référence et dialogue. In *Cinquième école d'été traitement des langues naturelles, Trégastel*, 1995.

[18] Marc Guyomard and Jacques Siroux. *The Structure of Multimodal Dialogue*, chapter Suggestive and Corrective Answers : a Single Mecanism, pages P. 361–374. Elsevier North-Holland, 1989.

[19] Marc Guyomard, Jaques Siroux, and Laurent Trilling. Le dialogueur : un intermédiaire entre l'utilisateur et l'application. In *Proceedings of the seminar Man-Machine Dialog by Voice, GRECO n ° 39 Communication parlée, CNRS, Nancy*, 1984.

[20] Philip J. Hayes. Anaphora for limited domain systems. In *Proceedings IJCAI*, pages 416–422, 1981.

[21] Peter A. Heeman and Graeme Hirst. Collaborating on referring expressions. *Computational Linguistics*, Volume 21, number 3 :351–382, 1995.

[22] Daniel Hernández. *Qualitative Representation of Spatial Knowledge*. Springer Berlin / Heidelberg, 1994.

[23] Jerry Hobbs. Pronoun resolution. Research Report 76-1, City College, City University of New York, 1976.

[24] Jerry Hobbs. Resolving pronoun references. *Lingua*, 44 :311–338, 1978.

[25] Hans Kamp and Uwe Reyle. *From discourse to logic*. Kluwer Academic Dordrecht, Boston, 1993.

[26] Frédéric Landragin. *Modélisation de la communication multimodale. Vers une formalisation de la pertinence*. PhD thesis, Université Henri Poincaré, Nancy 1, 2003.

[27] Shalom Lappin and Herbert J. Leass. An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20 :535Ũ561, 1994.

[28] Johan L'Hour, Olivier Boëffard, Jacques Siroux, Laurent Miclet, Francis Charpentier, and Thierry Moudenc. Doris, a multiagent/ip platform for multimodal dialogue applications. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, pages 3049–3052, Jeju Island, Korea, 2004.

[29] Hélène Manuélian. *Descriptions définies et démonstratives : analyses de corpus pour la génération de textes*. PhD thesis, Université Nancy 2, novembre 2003.

[30] Jean-Claude Martin. *Coopérations entre modalités et liage par synchronie dans les interfaces multimodales*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, mars 1995.

[31] Ruslan Mitkov. *Anaphora Resolution*. 0-582-32505-6. Pearson Education, 2002.

[32] Luis Pineda and Gabriela Garza. A model for multimodel reference resolution. *Computational Linguistics*, 26 (2) :139–193, 2000.

[33] Guillaume Pitel. *MICO : La notion de Construction Située pour un Modèle dŠInterprétation et de Traitement de la Référence pour le Dialogue Finalisé*. PhD thesis, Université Paris XI, 2004.

[34] Massimo Poesio, Tomonori Ishikawa, Sabine Schulte im Walde, and Renata Viera. Acquiring lexical knowledge for anaphora resolution. In *In Proceedings of the 3rd Conference on Language Resources and Evaluation (LREC)*, pages 1220–1224, 2002.

[35] Susanne Salmon-Alt. *Référence et dialogue finalisé : de la linguistique à un modèle opérationnel*. PhD thesis, Université de Nancy 1, 2001.

[36] Daniel Schang. *Représentation et interprétation de connaissances spatiales dans un système de dialogue homme-machine*. PhD thesis, Université Henri Poincaré, Nancy 1, 1997.

[37] Jacques Siroux. Le contexte dans les systèmes de dialogue personne-machine. In *Cinquième école d'été traitement des langues naturelles, Trégastel*, 1995.

[38] Masaru Tomita and Jaime G Carbonell. The universal parser architecture for knowledge-based machine translation. *IJCAI*, pages 718–721, 1987.

[39] Claude Vandeloise. *L'espace en français*. Editions du Seuil, Paris, 1986.

[40] Renata Vieira and Massimo Poesio. An empirically-based system for processing definite descriptions. *Computational Linguistics*, 26 (4) :539–593, 2000.

[41] Donald E. Walker, William H. Paxton, Jane J. Robinson, Gary G. Hendrix, Ben G. Deutsch, and Ann E. Robinson. Speech understanding system. Technical report, SRI, 1975.

[42] Terry Winograd. *Understanding natural language*. New York, Academic Press, 1972.

Table des matières